The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Nonhuman Evidence for PFOA Effects on Fetal Growth

Erica Koustas,[1] Juleen Lam,[1] Patrice Sutton,[2] Paula I. Johnson,[2] Dylan S. Atchley,[2] Saunak Sen,[3] Karen A. Robinson,[4,5,6] Daniel A. Axelrad,[7] and Tracey J. Woodruff[2]

[1]Oak Ridge Institute for Science and Education (ORISE) Postdoctoral Fellow, National Center for Environmental Economics, Office of Policy, U.S. Environmental Protection Agency, Washington, DC, USA; [2]Program on Reproductive Health and the Environment, University of California, San Francisco, Oakland, California, USA; [3]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA; [4]Department of Medicine, [5]Department of Epidemiology, and [6]Health Policy & Management, Johns Hopkins School of Medicine, Baltimore, Maryland, USA; [7]National Center for Environmental Economics, Office of Policy, U.S. Environmental Protection Agency, Washington, DC, USA

Background: In contrast to current methods of expert-based narrative review, the Navigation Guide is a systematic and transparent method for synthesizing environmental health research from multiple evidence streams. The Navigation Guide was developed to effectively and efficiently translate the available scientific evidence into timely prevention-oriented action.

Objectives: We applied the Navigation Guide systematic review method to answer the question "Does fetal developmental exposure to perfluorooctanoic acid (PFOA) or its salts affect fetal growth in animals?" and to rate the strength of the experimental animal evidence.

Methods: We conducted a comprehensive search of the literature, applied prespecified criteria to the search results to identify relevant studies, extracted data from studies, obtained additional information from study authors, conducted meta-analyses, and rated the overall quality and strength of the evidence.

Results: Twenty-one studies met the inclusion criteria. From the meta-analysis of eight mouse gavage data sets, we estimated that exposure of pregnant mice to increasing concentrations of PFOA was associated with a change in mean pup birth weight of –0.023 g (95% CI: –0.029, –0.016) per 1-unit increase in dose (milligrams per kilogram body weight per day). The evidence, consisting of 15 mammalian and 6 nonmammalian studies, was rated as "moderate" and "low" quality, respectively.

Conclusion: Based on this first application of the Navigation Guide methodology, we found sufficient evidence that fetal developmental exposure to PFOA reduces fetal growth in animals.

Citation: Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, Robinson KA, Axelrad DA, Woodruff TJ. 2014. The Navigation Guide—evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ Health Perspect 122:1015–1027; http://dx.doi.org/10.1289/ehp.1307177

Introduction

Background. In clinical research, systematic reviews have played a transformative role as a transparent, robust method for synthesizing the available evidence for incorporation into more efficient guidelines and recommendations related to medical interventions. Whereas systematic review methodology has been developed and tested in the clinical sciences for making evidence-based decisions for medical interventions, the methods are not fully transferable to environmental health science, largely because of their primary application to randomized controlled clinical trials, which are, for primarily ethical reasons, unavailable in environmental health. The Navigation Guide was developed to bridge this gap between clinical and environmental health sciences. The methodology provides the capacity to systematically and transparently evaluate the quality and strength of evidence from both human and nonhuman streams of evidence about the relationship between the environment and reproductive and developmental health (Woodruff and Sutton 2014; Woodruff et al. 2011a).

To test and refine the Navigation Guide systematic review methodology, we applied it to the evaluation of experimental animal evidence for the effects of exposure to the environmental contaminant perfluorooctanoic acid (PFOA) on fetal growth. The results of applying the method to the human evidence and integrating the animal and human data into an overarching strength of evidence rating are presented elsewhere (Johnson et al. 2014; Lam et al. 2014).

Rationale for selecting PFOA. Environmental exposures to the industrial chemical PFOA are widespread, and PFOA has been detected in the blood of > 95% of the U.S. population [Agency for Toxic Substances and Disease Registry (ATSDR) 2009; Centers for Disease Control and Prevention (CDC) 2009; Kato et al. 2011; U.S. Environmental

Address correspondence to E. Koustas, U.S. Environmental Protection Agency, 1200 Pennsylvania Ave. NW (1809T), Washington, DC 20460 USA. Telephone: (202) 566-1829. E-mail: koustas.erica@epa.gov

Supplemental Material is available online (http://dx.doi.org/10.1289/ehp.1307177).

We thank K. Guyton [U.S. Environmental Protection Agency (EPA)] for her contribution to developing the evidence evaluation methodologies and assistance with identifying relevant toxicological databases; A. Rooney (National Toxicology Program) for his toxicological expertise and for assistance with developing animal risk of bias criteria; and K. Thayer (National Toxicology Program), L. Bero [University of California, San Francisco (UCSF)], and L. Zeise (California Environmental Protection Agency) for assistance with developing risk of bias evaluation criteria, meta-analysis methodologies, and evidence evaluation methodologies. M. Macleod (University of Edinburgh) provided critical assistance in developing the animal-specific systematic review methodology. We thank H. Vesterinen (UCSF) for her thoughtful comments on this manuscript. C. Lau (U.S. EPA) served as a perfluorooctanoic acid (PFOA) subject matter expert. We thank T. Horvath (UCSF) for training and assistance with developing our search strategy. Finally, we acknowledge B. Abbott, J. Boberg, J. Butenhoff, S. Fenton, A. Hagenaars, J. Dai, J. DeWitt, C. Lau, G. Kennedy, S. Kennedy, J. O'Brien, N. Onishchenko, A. Pinkas, S. White, J. Yanai, and their laboratories for their kind and timely response to requests for additional information.

This research was supported, in part, by appointments of E.K. and J.L. to the Internship/Research Participation Program at the National Center for Environmental Economics, U.S. EPA, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. EPA. T.J.W. is a member of the Philip R. Lee Institute for Health Policy Studies. Financial support for development of the PFOA case study was through grants from the New York Community Trust and the U.S. EPA through a contract with Abt Associates (GAIA-0-6-UCSF 17288). For 2009–2013, support for the development and dissemination of the Navigation Guide methodology was provided by the Clarence Heller Foundation, the Passport Foundation, the Forsythia Foundation, the Johnson Family Foundation, the Heinz Endowments, the Fred Gellert Foundation, the Rose Foundation, Kaiser Permanente, the New York Community Trust, the Philip R. Lee Institute for Health Policy Studies, the Planned Parenthood Federation of America, the National Institute of Environmental Health Sciences (grants ES018135 and ES022841), and U.S. EPA STAR grants (RD83467801 and RD83543301).

The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the U.S. EPA. Further, the U.S. EPA does not endorse the purchase of any commercial products or services mentioned in the publication.

The authors declare they have no actual or potential competing financial interests.

Received: 4 June 2014; Accepted: 23 April 2014; Advance Publication: 25 June 2014; Final Publication: 1 October 2014.

Protection Agency (EPA) 2009] and in blood samples throughout the world (Kannan et al. 2004; U.S. EPA 2009). Voluntary efforts by eight major manufacturers of PFOA to eliminate global emissions and product content by the end of 2015 are ongoing, and significant progress has been made for both U.S. and non-U.S. operations (U.S. EPA 2008, 2012, 2013a). However, PFOA can remain in the environment, and with a half-life in humans of approximately 3.5 years (Olsen et al. 2007), the chemical will persist in people for years to come (U.S. EPA 2013b, 2014).

Fetal exposure to PFOA may be widespread because the chemical is ubiquitous in the blood of pregnant women and women of child-bearing age and in cord blood (Apelberg et al. 2007a; Calafat et al. 2007; Fei et al. 2007; Midasch et al. 2007; Mondal et al. 2012; Monroy et al. 2008; Woodruff et al. 2011b). The association between PFOA and fetal growth reported in individual human studies has been inconsistent, with some reporting statistically significant associations between prenatal exposure to PFOA and restricted fetal growth (Apelberg et al. 2007b; Fei et al. 2007, 2008) and others reporting no or nonstatistically significant associations (Hamm et al. 2010; Monroy et al. 2008; Washino et al. 2009). The animal literature also includes reports of inconsistent associations between PFOA and fetal growth, including findings of reduced birth weight following prenatal exposure to PFOA in rodents (Butenhoff et al. 2004; Hines et al. 2009; Lau et al. 2004, 2006).

Ubiquitous exposure to a chemical that lacks evidence of nontoxicity is a potential public health concern; moreover, PFOA has been associated with adverse impacts on the quality and duration of the gestation period—one of the most important indicators of an infant's health and survival (Gluckman and Hanson 2006; Institute of Medicine 2007). Given the potential concern for an adverse developmental health outcome of public health importance and the availability of data, we selected PFOA to test and refine the Navigation Guide method.

Methods

We assembled a review team to include experts in the fields of risk assessment, environmental health, epidemiology, biology, systematic review, and toxicology to develop a protocol that covered the first three steps of the Navigation Guide systematic review method: *1*) Specify the study question; *2*) select the evidence; and *3*) rate the quality and strength of the evidence (Koustas et al. 2013). Each of the steps of the Navigation Guide method described below involves application of standardized and transparent documentation, including expert judgment. Additional information regarding the Navigation Guide methodology is available elsewhere (Woodruff and Sutton 2014).

Step 1. Specify the Study Question

Our objective was to answer the question: "Does fetal developmental exposure to PFOA or its salts affect fetal growth in animals?" PICO (participants, interventions, comparator, and outcomes) is an aid used to formulate an answerable question in a systematic review and to provide more specific information about the scope of the review (O'Connor et al. 2011). Because we were evaluating environmental exposures, we used the acronym PECO (i.e., participants, exposure, comparator, and outcomes).

Participants. Animals were from nonhuman species studied during the reproductive/developmental time period (before and/or during pregnancy for females or during development for embryos).

Exposure. Exposure included one or more oral, subcutaneous, or other treatment(s) of any dosage of PFOA (CAS# 335-67-1) or its salts during the time before pregnancy and/or during pregnancy for females or directly to embryos.

Comparator. Experimental animals receiving different doses of PFOA or vehicle-only treatment were used for comparisons.

Outcomes. The outcomes examined for mammalian species were fetal weight near term (e.g., embryonic day 18 for mice and embryonic day 21 for rats) or at birth, and/or other measures of size near term or at birth, such as length. For nonmammalian species, outcomes were weight and/or other measures of size in late stages of embryonic development.

Step 2. Select the Evidence

Search methods. Our search was developed by analyzing the Medical Subject Headings (MeSH) and other terms from the title and abstract text of a group of seven papers known to us, judged to be relevant to our study question, and which represented different journals and years of publication (Abbott et al. 2007; Butenhoff et al. 2004; Hines et al. 2009; Lau et al. 2006; Staples et al. 1984; White et al. 2007, 2009). A list of common and unique terms was compiled and incorporated into a search strategy to address the exposure (PFOA) and outcomes of interest (reproductive/developmental toxicity), as defined in the PECO statement (see Supplemental Material, Tables S1, S2). To develop search terms to retrieve experimental animal studies, we adapted a search filter developed by Hooijmans et al. (2010b).

We searched PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Web of Science (http://apps.webofknowledge.com/) on 3 February 2012. Using PFOA terms, we searched 35 toxicological databases between 23 January and 6 February 2012. Our search was not limited by language or publication date. We hand searched the reference list of all included studies and searched for publications citing the included studies. We also consulted with a subject matter expert (C. Lau, U.S. EPA).

Data collection and management. We imported or manually entered all retrieved records into EndNote X4 reference management software (http://thomsonreuters.com/endnote/), and each record was assigned a source identification number, which was used to track individual studies throughout the course of the review. Two authors (E.K. and J.L.) independently screened the titles and abstracts of each record retrieved to identify those meeting our inclusion criteria using DistillerSR (Evidence Partners; http://www.systematic-review.net). We developed inclusion and exclusion criteria based on our PECO statement. All studies that compared experimental animals exposed to one or more doses of PFOA during reproductive or developmental periods with untreated control experimental animals were eligible for inclusion. We excluded studies if one or more of the following criteria were met: the article did not contain original data (i.e., review article); study subjects were not animals; PFOA was not administered to study subjects; and PFOA was not administered during the reproductive/developmental time period. Two authors (E.K. and J.L.) assessed the full text of studies that could not be excluded based on screening of the title and abstract. Potentially relevant non-English articles were translated to determine eligibility. To provide quality control, a third author (P.I.J.) screened the title and abstract of 5% of the search results and 5% or five articles—whichever was greater—of the search results eligible for full text. We considered studies that described more than one experiment or outcome measure as separate data sets.

Two authors (E.K. and J.L.) independently extracted data relating to study characteristics and outcome measures from all included articles into a Microsoft Access 2007 database. The list of extracted study characteristics was based on a compilation of previously published checklists and criteria (Guyatt et al. 2011; Higgins and Deeks 2011; Hooijmans et al. 2010a; Kilkenny et al. 2010).

One author (E.K.) performed data entry of the raw outcome data using Microsoft Excel 2007, and a second author (D.S.A.) verified all values. We contacted study authors when additional information was required for performing statistical analysis and/or analysis of the full data set. For example, we requested

numerical estimates associated with figures presented in published articles, numbers of animals allocated to various test groups, and raw data values. In some cases, fetal growth data were not presented in the published study because the outcome was not of primary interest to the study authors. If there was a reason to believe the study authors may have measured fetal growth, we contacted them to obtain any data they may have collected during the course of the study. We also contacted all study authors to inform them of our systematic review and to verify values used in both the meta-analysis and analysis of the full data set.

Statistical analyses. Two authors (E.K. and J.L.) assessed study characteristics from all included articles for comparability (i.e., study features and biological heterogeneity) to determine which studies were suitable for meta-analysis. We consulted experts in the field of PFOA toxicity, toxicological study design, or human/animal toxicity reviews to develop these characteristics and their associated heterogeneity concerns beforehand. For example, we considered the differences in PFOA clearance rates between female mice (approximately 17 days) and rats (2–4 hr) as a potential biological heterogeneity concern (Lau et al. 2007).

From the assessment of specified characteristics, we determined that only a subset of data was combinable in a meta-analysis. This subset of seven studies (eight data sets) had the following characteristics:

- Species: mouse
- Route of exposure: gavage
- Method of outcome measurement: weight
- Time point of outcome measurement: at or soon after birth.

We used the mean (± SE) pup body weight at birth from each of the eight data sets for all PFOA doses < 5 mg/kg body weight (mg/kg BW)/day. We limited the dose range in order to focus on effects at lower tested doses and to minimize adverse impacts from responses at higher doses (such as litter loss) on the overall estimate. We used a two-step modeling approach. In the first step, we analyzed each data set separately using a linear mixed-effects model, and we obtained a slope estimate of the dose–response effect (and the associated SE). In the second step, we combined the slope and SE estimate from each data set using a random-effects model.

The result was an estimate of the overall mean change in body weight per offspring for a 1-unit increase in mg/kg BW/day dose, accounting for within- and between-study variability. We used the programming environment R, version 2.13.1 (R Development Core Team; http://www.R-project.org/) and its standard packages. We used the metafor package in R (Viechtbauer 2010) for conducting our random effects meta-analysis.

To visually assess the possibility of publication bias in a meta-analysis, we considered producing a funnel plot of the estimated effects. However, tests for funnel plot asymmetry are not recommended when there are < 10 studies because test power is usually too low to distinguish chance from real asymmetry (Sterne et al. 2011a). Because our meta-analysis was limited to 7 studies (8 data sets), we did not produce a funnel plot.

Statistical heterogeneity assessment. We sought to assess whether differences in estimated effect sizes among studies were consistent with random variation versus nonrandom heterogeneity among the studies. We estimated the between-study variance component and tested the null hypothesis that the between-study variability was absent using Cochran's Q statistic. The test statistic follows a chi-square distribution with $n - 1$ degrees of freedom, where n is the number of studies. We considered $p \leq 0.05$ statistically significant. We also calculated the I^2 statistic, which estimates the percentage of variation across studies due to heterogeneity rather than chance (Higgins et al. 2003), and used the Cochrane Collaboration's guidelines to interpret the statistic, where a value of > 50% may indicate substantial heterogeneity (Deeks et al. 2011). To assess the overall impact of existing study heterogeneity on the meta-analysis, we considered the magnitude/direction of effect estimates, the I^2 statistic, and the p-value from the Cochran's Q test.

Sensitivity analysis. We performed sensitivity analyses using subgroup analyses based on characteristics described above that were used to determine comparability across studies for the meta-analysis. To evaluate the influence of each individual study on the main meta-analysis results and assist in identifying any study characteristics that might be influential in the final results, we performed a sensitivity analysis by removing one data set at a time from the meta-analysis.

Analysis of the full data set. We analyzed all included animal studies identified via our search and exclusion/inclusion assessment to assess the totality of all available animal evidence. This was done to maximize use of all data, in addition to those determined appropriate to combine in the meta-analysis. To assess results from the full data set, we calculated the percentage change in outcome (weight or length) compared with the control group for each tested dose group for each of the data sets and used these values to create scatter plots. Two of the nonmammalian studies reported outcome measurements at multiple time points during larval development (Spachmo and Arukwe 2012; Wang et al. 2010). We selected the outcome measurement reported at the latest time point during the larval stage, based on justification that this allowed for consideration of maximal larval growth. For each study, we used the mean and SE estimates reported by authors to calculate a 95% confidence interval (CI) for the difference in means comparing each treatment group with the control group. We interpreted a 95% CI that overlapped zero as indicating no statistically significant difference between the mean weight in that treatment group with the mean weight in the control group.

Step 3. Rate the Quality and Strength of the Evidence

To rate the evidence, we *a*) determined the risk of bias for individual studies based on seven domains; *b*) rated the overall quality across all studies in the body of evidence based on five factors, including risk of bias; and *c*) rated the overall strength of evidence across all studies in the body of evidence based on four considerations, including the quality of the body of evidence (Figure 1).

Assessment of risk of bias. Two authors (E.K. and J.L.) assessed risk of bias defined as characteristics of a study that can introduce a systematic error in the magnitude or direction of the results of the study (Higgins et al. 2011) for included studies based on seven risk of bias domains using modified terminology and concepts in the Cochrane Collaboration's Tool for Assessing Risk of Bias. Informed by empirical data from meta-analyses conducted on pharmacological treatments (Roseman et al. 2011), we considered funding source and reported conflicts of interest to be potential sources of bias. We did not ask study authors for additional information to inform our risk of bias determinations. However, if study authors mentioned study design details in their responses to our requests for data, we considered the information while evaluating risk of bias. See Table 1 for a summary of the risk of bias domains assessed for each included study (see also Supplemental Material, "Instructions for making risk of bias determinations").

Rate the quality and strength of the body of evidence. Upon completion of the data analysis, each of the nine review authors compared the results of the systematic review to the Navigation Guide factors and considerations for rating the quality and strength of the nonhuman evidence. The Navigation Guide rating method (Woodruff et al. 2011a) was applied according to explicit written directions (Koustas et al. 2013). Because of fundamental biological differences between mammalian and nonmammalian model systems, we evaluated the mammalian and nonmammalian studies as separate bodies of evidence.

The possible ratings for the overall quality of the body of evidence were "high," "moderate," and "low." These quality ratings were determined by assigning an initial rating according to the type of study, and then downgrading the rating if factors that decrease the quality level of the studies were present. The initial quality rating assigned to both the mammalian and nonmammalian bodies of evidence was "high," comparable to the rating assigned to human experimental studies [i.e., randomized controlled trials (RCTs)] in systematic review methods used in the clinical sciences. An initial "high" quality rating for experimental animal studies was supported by the level of study control exercised in such studies and the limited heterogeneity within an experimental animal study population. This is also consistent with Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidelines for clinical evidence that consider randomization a key determinant of a "high" grade (Guyatt et al. 2011). Upgrades to the quality rating for experimental animal data were not considered because the initial quality level was "high."

The overall body of evidence was evaluated for downgrading based on the presence of five factors (Figure 1):

1. Risk of bias across studies: a substantial risk of bias existed across the body of evidence.
2. Indirectness: Evidence was not directly comparable to the question of interest (i.e., population, exposure, comparator, and/or outcome). Beforehand, we decided not to downgrade experimental animal studies for indirectness because studies find that humans are as sensitive or more sensitive to chemical exposures than animals, strengthening the applicability of findings from experimental animal studies to human health outcomes (Kimmel et al. 1984; U.S. EPA 1996). However, in applying GRADE principles to the Navigation Guide, evidence would be rated down if the animal model was determined to be biologically inappropriate for the health outcome under study.
3. Inconsistency: Estimates of effect were widely different (heterogeneity or variability in results).
4. Imprecision: Studies had few participants and few events (wide CIs).
5. Publication bias: Studies were missing from the body of evidence, resulting in an overestimate or underestimate of true effects from exposure.

According to GRADE, these five factors address nearly all issues that bear on the quality of evidence (Balshem et al. 2011). Each of the nine review authors reviewed the body of evidence and applied their expert judgment to independently and transparently grade the quality of evidence based on the presence of the five objective factors using detailed instructions (Koustas et al. 2013). Possible ratings were 0 (no change from "high" quality), –1 (one-level downgrade to "moderate" quality) or –2 (two-level downgrade to "low" quality). Consistent with GRADE's approach to evaluating risk of bias across studies (Guyatt et al. 2011), authors

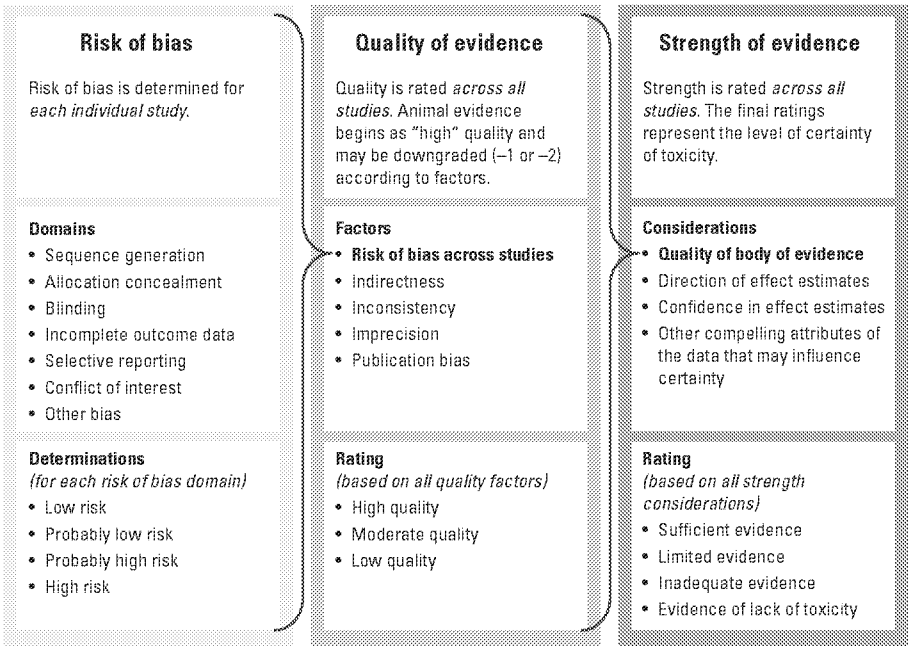

Figure 1. Flowchart for evaluating risk of bias, quality of evidence, and strength of evidence.

Table 1. Tool for assessing risk of bias.

Domain	Criteria for low risk of bias rating	Examples of factors considered in assessment
Sequence generation	Study authors reported the use of a random component in the sequence generation process.	Use of a random component, such as a random number table or computer random number generation; statement by study author that animals were randomly allocated.
Allocation concealment	Study authors reported that study personnel could not foresee which animals were allocated to the various experimental groups.	Use of sequentially numbered cages or animals.
Blinding	Study authors reported that personnel and outcome assessors were adequately prevented from knowledge of the allocated exposures during the study.	Use of masked identifiers in the study and for outcome assessment.
Incomplete outcome data	Study authors reported when and why participants left the study.	The number of animals allocated to experimental groups was reported and/or adequate follow up of dams and offspring (for mammalian studies) was carried out; the number of organisms allocated to experimental groups was reported and/or organisms were adequately followed up after exposure (for nonmammalian studies).
Selective reporting	The study's prespecified outcomes that are of interest in the review were reported in a prespecified way.	The number of animals or organisms analyzed for outcomes of interest was reported, or study authors provided additional data; study methods matched study results for outcomes of interest.
Conflict of interest	The study was free of support from a company, study author, or other entity having a financial interest in the exposures of interest in the review.	The study was funded or conducted by companies with a financial interest in PFOA; companies provided services to assist in the completion of the study, evaluate the data, or write the manuscript; or the publication or report included a declaration of conflicts of interest.
Other bias	Study appears to be free of other sources of bias.	Other potential sources of bias related to the specific study design.

were instructed to be conservative in making judgments to downgrade the evidence for all factors (i.e., high confidence of substantial concerns with the body of evidence before rating down). Authors reviewed the body of evidence as a way to initiate the group discussion and gather all perspectives for consideration. After independently evaluating the quality of the evidence, all authors discussed their evaluations. The discussion between coauthors was extensive and iterative and was carried out over several meetings until a consensus was reached. These collective decisions did not involve a "majority vote" or other tallying of perspectives. It was prespecified that discrepancies between review authors that could not be resolved through consensus would be resolved by the senior author (T.J.W.). However, for this case study, review authors were able to agree on a collective consensus for each rating and the arbiter was not necessary. The rationale for each collective decision on each of the five factors was recorded.

In systematic reviews in the clinical sciences, rating the quality of evidence is the final step because only one stream of evidence is considered in a decision. However, given that our purpose was to ultimately integrate the strength of multiple streams of evidence used in environmental health decision making (i.e., toxicology and epidemiology) leading to a concise "bottom line" statement about a chemical's toxicity that brings all of the relevant evidence to bear, the Navigation Guide systematic review method specifies an additional step: moving from quality of evidence to strength of evidence.

We rated the overall strength of the evidence based on a combination of four considerations: *a*) quality of the body of evidence, *b*) direction of effect estimates, *c*) confidence in effect estimates (likelihood that a new study would change our conclusion), and *d*) other compelling attributes of the data that may influence certainty (Figure 1). The results of rating the strength of the nonhuman evidence were compared with the definitions specified in the Navigation Guide for "sufficient" evidence of toxicity; "limited" evidence of toxicity; "inadequate" evidence of toxicity; or "evidence of lack of toxicity" (Table 2), which were based on criteria in use by the International Agency for Research on Cancer (IARC 2006) and the U.S. EPA (1991, 1996). The procedure for rating the strength of the evidence was similar to rating the quality of evidence: All review authors independently evaluated the strength of the evidence according to the same four considerations, and then they compared their evaluations, resolved any discrepancies through discussion, and recorded the rationale for every collective decision.

Results

Included Studies

Of the 2,049 unique records we identified (see Supplemental Material, Table S3 for the total number of hits retrieved from each database), 1,982 were excluded through title and abstract screening and 46 articles were excluded during full-text review, resulting in 21 studies describing 32 data sets included in the review (Figure 2). A summary of mammalian and nonmammalian study characteristics are provided in Tables 3 and 4, respectively. Detailed characteristics of each mammalian and nonmammalian study are provided in Supplemental Material, Tables S4–18 and Tables S19–S24, respectively. Various details of outcome data and study design characteristics necessary for data analysis were missing from all 21 articles. In some cases, published articles did not include details needed for our analysis, such as numerical outcome measurements or data on fetal growth, if this was not a primary outcome of interest for study authors. In other cases basic information, such as allocation numbers or the number of animals weighed to obtain given outcome values, was missing. Our efforts to contact study authors resulted in obtaining additional data for 18 of the 21 included studies, along with raw data in many instances (see Supplemental Material, Tables S4–S24).

Populations. Of the 21 studies, 15 were conducted on mammalian species (11 mouse and 4 rat) and 6 studies were conducted on nonmammalian species (3 chicken, 1 fruit fly, 1 zebrafish, and 1 salmon) (Tables 3 and 4).

Mammalian exposures. For all 15 mammalian studies, pregnant female dams were exposed to PFOA, and fetal growth was measured in the resulting progeny (Table 3). The primary route of exposure was oral gavage (13 studies), but some studies also evaluated exposures via inhalation, food, and water. Most of the mammalian studies (12) exposed dams to the ammonium salt form of PFOA (CAS# 3825-26-1), 1 study exposed dams to the free acid form (CAS# 335-67-1), and 2 studies did not specify the form used for exposure. The dose range tested varied widely across studies, ranging from 0.01 to 100 mg/kg BW/day. Inhalation study doses ranged between 0.1 and 25 mg/m^3. The number of PFOA doses administered per study ranged from one to six. Although dams in all studies were exposed to PFOA at some point during their pregnancy, the window of exposure varied across studies

Table 2. Strength of evidence definitions for nonhuman studies.[a]

Strength rating	Definition
Sufficient evidence of toxicity	A positive relationship is observed between exposure and adverse outcome in multiple studies or a single appropriate study in a single species.[b] The available evidence includes results from one or more well-designed, well-conducted studies, and the conclusion is unlikely to be strongly affected by the results of future studies.[c]
Limited evidence of toxicity	The data suggest a positive relationship between exposure and adverse outcome, but there are important limitations in the quality of the body of evidence. Confidence in the relationship is constrained by factors such as the number, size, or quality of individual studies, or inconsistency of findings across individual studies.[c] As more information becomes available, the observed effect could change, and this change may be large enough to alter the conclusion.
Inadequate evidence of toxicity	The available evidence is insufficient to assess effects of the exposure. Evidence is insufficient because of the limited number or size of studies, low quality of individual studies, or inconsistency of findings across individual studies. More information may allow an assessment of effects.
Evidence of lack of toxicity	Data on an adequate array of end points from more than one study with at least two species showed no adverse effects at doses that were minimally toxic in terms of inducing an adverse effect. Information on pharmacokinetics, mechanisms, or known properties of the chemical class may also strengthen the evidence.[d] The conclusion is limited to the species, age at exposure, and/or other conditions and levels of exposure studied, and is unlikely to be strongly affected by the results of future studies.[c]

[a]The Navigation Guide rates the quality and strength of evidence of human and nonhuman evidence streams separately as "sufficient," "limited," "inadequate," or "evidence of lack of toxicity," and then these two ratings are combined to produce one of five possible statements about the overall strength of the evidence of a chemical's reproductive/developmental toxicity. The methodology is adapted from the criteria used by IARC to categorize the carcinogenicity of substances (IARC 2006) except as noted. [b]IARC's criteria for sufficient evidence of carcinogenicity in animals requires multiple positive results (species, studies, sexes) (IARC 2006). The Navigation Guide integrates the U.S. EPA's minimum criteria for animal data for a reproductive or developmental hazard (i.e., data demonstrating an adverse reproductive effect in a single appropriate, well-executed study in a single test species) (U.S. EPA 1996). The Navigation Guide also incorporates the U.S. EPA's "sufficient evidence category," which includes data that "collectively provide enough information to judge whether or not a reproductive hazard exists within the context of effect as well as dose, duration, timing, and route of exposure. This category may include both human and experimental animal evidence" (U.S. EPA 1996). The U.S. EPA statement for developmental hazards is slightly different but includes the same relevant information regarding dose, duration, timing, and so on (U.S. EPA 1991). [c]Language for the definitions of the rating categories were adapted from descriptions of levels of certainty provided by the U.S. Preventive Services Task Force Levels of Certainty Regarding Net Benefit (Sawaya et al. 2007). [d]Based on minimum data requirements according to U.S. EPA guidelines for assessing reproductive toxicity (U.S. EPA 1996).

from a single gavage exposure on a single day of pregnancy to exposure prior to conception that continued throughout pregnancy.

Mammalian comparators. Eleven gavage studies used water as a vehicle control and two used corn oil (Table 3). The inhalation study utilized three control groups: in-house air only, in-house air pair-fed 10 mg/m^3 PFOA, and in-house air pair-fed 25 mg/m^3 PFOA (Staples et al. 1984). The PFOA-treated food study (Onishchenko et al. 2011) used ethanol-treated food as a control, and the PFOA-treated water study (Hu et al. 2010) used untreated water as a control. Other than PFOA exposure, all control groups were treated similarly to dose groups for each data set.

Mammalian outcomes. Body weight was used as the outcome measure for all 15 mammalian studies (Table 3). Because pregnant dams were exposed to PFOA for all mammalian studies, the litter was used as the statistical unit; the total number analyzed across studies ranged from 8 to 183 litters.

The time point of weight measurement varied between fetal time points near term, typically gestation day (GD) 18 for mice and GD21 for rats, to at or near the time of birth, typically postnatal day (PND) 0 to PND2. The methods used to monitor parturition varied widely across birth weight studies, from constant monitoring to daily cage checks. PND1 was defined as either the day of birth or the day after birth.

The method of weight measurement varied across studies as well, from weighing offspring individually, grouped by litter, or grouped by sex, to weighing a subset of offspring from each litter. Offspring survival was statistically significantly reduced (based on the alpha level specified by study authors, generally < 0.05 or ≤ 0.05) at exposure to doses > 5 mg/kg BW/day in five studies; one study did not provide statistics or comment on litter sizes at birth.

Mammalian risk of bias assessment. On the basis of our risk of bias assessment, we concluded that the majority of studies had "probably high" risk of bias for allocation concealment and blinding, and "probably low" risk of bias for incomplete outcome data and selective reporting. Ratings for sequence generation and conflict of interest were mixed across studies, and ranged from low to high risk of bias. All studies had low risk of bias for the "other bias" domain (Figure 3A,B). See Supplemental Material, Tables S25–S39, for details on the risk of bias results for each mammalian study.

Nonmammalian exposures. Developing embryos were directly exposed to PFOA in all six nonmammalian studies (Table 4). Routes of administration varied based on test species: injection of PFOA solution into eggs for chickens, immersion of eggs in PFOA solution for zebrafish and salmon, and PFOA-treated food for fruit flies. One study exposed organisms to the ammonium salt form of PFOA (CAS#3825-26-1), two studies exposed organisms to the free acid form of PFOA (CAS#335-67-1), and three studies did not specify the form of PFOA. The dose ranges across studies varied based on animal species tested: chicken (0.01–10 mg/kg egg), zebrafish (15–250 mg/L water), fruit fly (100–500 μM in food), and salmon (100 μg/L water). The number of PFOA doses administered per study ranged from one to eight.

In all nonmammalian studies, embryos were exposed during development, and the time period of exposure varied based on species. For the chicken studies, a single injection of PFOA was administered to eggs on incubation day 0; for zebrafish studies, eggs were exposed from 60 min after spawning to 120 hours post fertilization (hpf); for salmon studies, eggs and larvae were exposed to PFOA-containing water for 48 days; for fruit fly studies, female flies were placed in vials with PFOA-containing food and allowed to lay eggs for a period of 2 hr, and eggs were allowed to hatch and develop through 110 hr after egg laying (ael) or to the white pupae stage, depending on data set.

Nonmammalian comparators. Chicken studies used saline, dimethyl sulfoxide (DMSO), or sunflower oil as vehicle controls, and some studies included an uninjected control (Table 4). The zebrafish study used water as a vehicle control, the fruit fly study used untreated food as a vehicle control, and the salmon study used water with carrier solvent (methanol) as a vehicle control. Besides PFOA exposure, all control groups were treated similarly to dose groups for each data set.

Nonmammalian outcomes. Relevant outcome measures varied across nonmammalian studies and included length, weight, and larval volume (calculated from measurements of length and diameter) (Table 4). Because embryos were directly exposed to PFOA in the nonmammalian model systems, the embryo was used as the unit of statistical analysis, and the total number of embryos analyzed across studies varied between 37 and 378.

The time points of outcome measurement varied: from shortly before time of hatching, shortly after hatching, and multiple time points during larval development.

PFOA exposure delayed hatching and larval emergence in the zebrafish and fruit fly studies and induced mortality in the zebrafish study and in one chicken study. Pipping success (i.e., when a chick breaks its shell) and the developmental stage at embryo death were unaffected by PFOA exposure in one chicken study, whereas in a second chicken study,

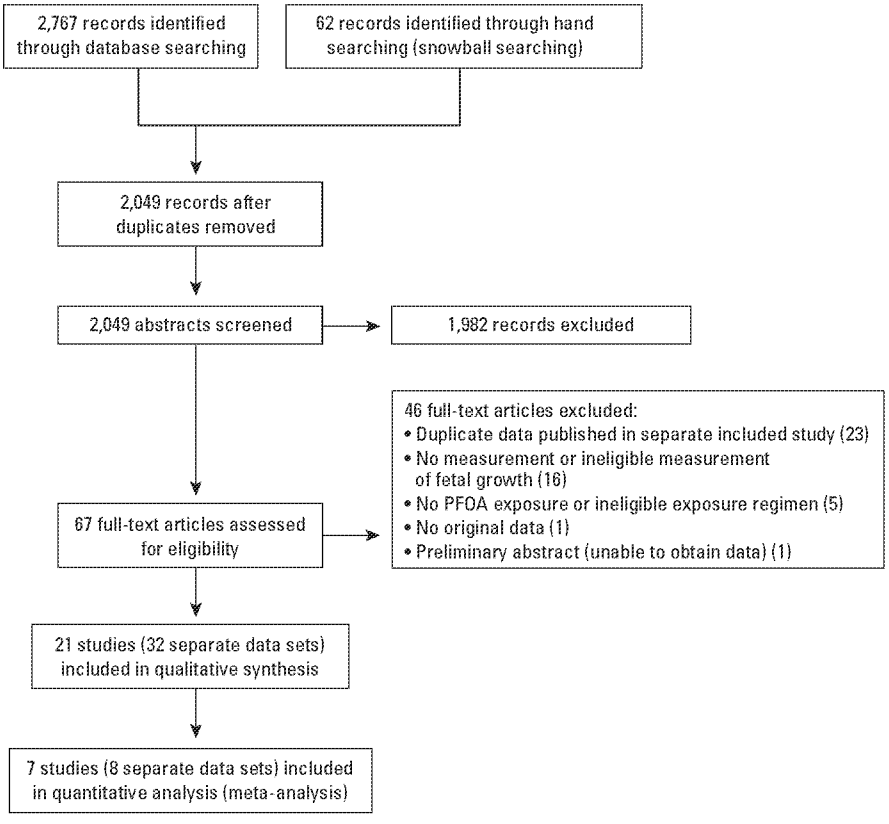

Figure 2. Flowchart of the study-selection process.

Table 3. Summary of mammalian study characteristics.

Source [source ID]	Species	Time point of outcome measurement	Outcome measure	Route of exposure	Period of exposure	PFOA dose range (mg/kg BW/day)[a]	No. of doses administered[b]	No. of litters	Reason(s) excluded from meta-analysis
Studies used in meta-analysis									
Hines et al. 2009 [260]	Mouse	Birth	Weight	Gavage	GDs 1–17	0.01–5	5	75	NA
White et al. 2009 [312]	Mouse	Birth	Weight	Gavage	GDs 8–17	5	1	8	NA
Abbott et al. 2007 [528]	Mouse	Birth	Weight	Gavage	GDs 1–17	0.1–1	4	58	NA
White et al. 2007 [566]	Mouse	Birth	Weight	Gavage	GDs 1–17 GDs 8–17 GDs 12–17	5	1	37	NA
Wolf et al. 2007 [571]	Mouse	Birth	Weight	Gavage	GDs 1–17	3–5	2	87	NA
Wolf et al. 2007 [571]	Mouse	Birth	Weight	Gavage	GDs 7–17 GDs 10–17 GDs 13–17 GDs 15–17	5–20	2	56	NA
Lau et al. 2006 [635][c]	Mouse	Birth	Weight	Gavage	GDs 1–17	1–20	5	103	NA
White et al. 2011 [3862]	Mouse	Birth	Weight	Gavage	GDs 1–17	1–5	2	60	NA
Studies not used in meta-analysis									
Hu et al. 2010 [68]	Mouse	Birth	Weight	Drinking water	GDs 6–17	0.05–1	2	30	Incomparable route of exposure
Yahia et al. 2010 [103]	Mouse	Fetal	Weight	Gavage	GDs 0–17	1–10	3	29	Incomparable time point of outcome measurement
Yahia et al. 2010 [103]	Mouse	Birth	Weight	Gavage	GDs 0–18	1–10	3	20	Time point of birth weight measurement was not specified
Fenton et al. 2009 [264]	Mouse	Fetal	Weight	Gavage	GD17	0.1–5	3	19	Incomparable time point of outcome measurement
Fenton et al. 2009 [264]	Mouse	Birth	Weight	Gavage	GD17	0.1–5	3	19	Dams were exposed for only 1 day of pregnancy
Lau et al. 2006 [635][c]	Mouse	Fetal	Weight	Gavage	GDs 1–17	1–40	6	183	Incomparable time point of outcome measurement
Hinderliter et al. 2005 [711][d]	Rat	Birth	Weight	Gavage	GDs 4–21	3–30	3	20	Incomparable species
Staples et al. 1984 [1871]	Rat	Fetal	Weight	Gavage	GDs 6–15	100	1	46	Incomparable species and time point of outcome measurement
Staples et al. 1984 [1871]	Rat	Fetal	Weight	Inhalation	GDs 6–15	0.1–25 mg/m^3	4[e]	103	Incomparable species, route of exposure, and time point of outcome measurement
Staples et al. 1984 [1871]	Rat	Birth	Weight	Gavage	GDs 6–15	100	1	21	Incomparable species
Staples et al. 1984 [1871]	Rat	Birth	Weight	Inhalation	GDs 6–15	0.1–25 mg/m^3	4	54	Incomparable species and route of exposure
Boberg et al. 2008 [3061]	Rat	Fetal	Weight	Gavage	GDs 7–20/21	20	1	11	Incomparable species and time point of outcome measurement
Onishchenko et al. 2011 [3610]	Mouse	Birth	Weight	Food	GDs 1–20	0.3	1	15	Incomparable route of exposure
York 2002 [5122][f]	Rat	Birth	Weight	Gavage	70 days prior to breeding throughout lactation	1–30	4	141	Incomparable species

GD, gestation day.
[a]Unless otherwise specified; the dose range is limited to those doses for which dams were analyzed. [b]Excludes control groups; study used one control group unless otherwise specified. [c]Lau (2006) is listed two times (birth weight data were included in meta-analysis; fetal data were excluded from meta-analysis). [d]Hinderliter (2005) is a peer-reviewed publication; the author provided an industry report with detailed data (Mylchreest 2003). [e]Included three control groups. [f]York (2002) is an industry report; the search also identified peer-reviewed journal publications describing findings from the report (Butenhoff et al. 2004; York et al. 2010), but these journal publications were excluded as duplicates because the report provided raw data.

Table 4. Summary of nonmammalian study characteristics.

Source [source ID]	Species	Time point(s) of outcome measurement	Outcome measure	Route of exposure	Period of exposure	PFOA dose range	No. of doses administered[a]	No. of offspring
Hagenaars et al. 2011 [59]	Zebrafish	120 hpf (posthatching)	Length	Egg immersion	Spawning, 120 hpf	15–250 mg/L	8	292
Wang et al. 2010 [86]	Fruit fly	30, 48, 72, 96, and 110 ael (larval stages)	Length[b]	Food	Egg laying, 110 ael	100–500 μM	2	378
Wang et al. 2010 [86]	Fruit fly	Pupae	Weight	Food	Egg laying, white pupae stage	100–500 μM	2	98
Pinkas et al. 2010 [187]	Chicken	Hatchling	Weight	Egg injection	Single treatment at incubation day 0	5–10 mg/kg egg	2	52
O'Brien et al. 2009 [236]	Chicken	Embryo at pipping star or day 22, whichever came first	Weight	Egg injection	Single treatment at incubation day 0	0.01–10 mg/kg egg	4[c]	37
Jiang et al. 2012 [3926]	Chicken	Embryonic day 19	Yolk-free body weight	Egg injection	Single treatment at incubation day 0	0.5–2 mg/kg egg	2[c]	40
Jiang et al. 2012 [3926]	Chicken	16–24 hr posthatching	Yolk-free body weight and crown to rump length	Egg injection	Single treatment at incubation day 0	0.5–2 mg/kg egg	2[c]	68
Spachmo and Arukwe 2012 [3932]	Salmon	Study days 21, 35, 49, 56 (larval stages posthatching)	Length and dry weight	Egg immersion	Egg stage, day 48	100 μg/L	1	80

Abbreviations: ael, hours after egg laying; hpf, hours postfertilization.
[a]Excludes control groups; study used one control group unless otherwise specified. [b]Length measurements provided by study author (used to calculate volume outcome reported in study). [c]Included two control groups.

embryonic mortality was increased but hatchling mortality and hatching success were not affected. The salmon study did not provide details on larval survival rates.

Nonmammalian risk of bias assessment. On the basis of our risk of bias assessment, we found that the majority of studies had probably high risk of bias for sequence generation, allocation concealment, and blinding, and probably low risk of bias for selective reporting. Ratings for incomplete outcome data were mixed across studies, and ranged from low to high risk of bias. Finally, all studies had probably low or low risk of bias for conflict of interest and low risk of bias for the "other bias" domain (Figure 3A,B). See Supplemental Material, Tables S40–S45, for details of the risk of bias results for each nonmammalian study.

Impact of PFOA on Fetal Growth

Analysis. Across the eight data sets determined to be combinable in the meta-analysis, gavage exposure of pregnant mice to increasing concentrations of PFOA was associated with a decrease in birth weight. The combined estimate from the meta-analysis was a change in mean pup birth weight of –0.023 g (95% CI: –0.029, –0.016) per 1-unit increase in dose (mg/kg BW/day) (Figure 4). The I^2 test statistic was calculated to be 0%, indicating no observed heterogeneity between studies that could not be explained by chance; this conclusion was further supported by the Q statistic, which produced a nonsignificant p-value of 0.73.

From the sensitivity analysis, where we removed one data set at a time, we found relatively small changes in the effect estimate, with a maximum of 9% change in the meta-analysis estimate (from –0.023 to –0.021) when the data set from White et al. (2011) was removed (data not shown). Figure 4 shows that the study of White et al. (2011) resulted in the largest estimate of decreased birth weight among those studies weighted more heavily in the meta-analysis (indicated by the larger size of the mean symbol), so it is not surprising that the removal of this study would have the largest effect on the meta-analysis estimate, and in particular

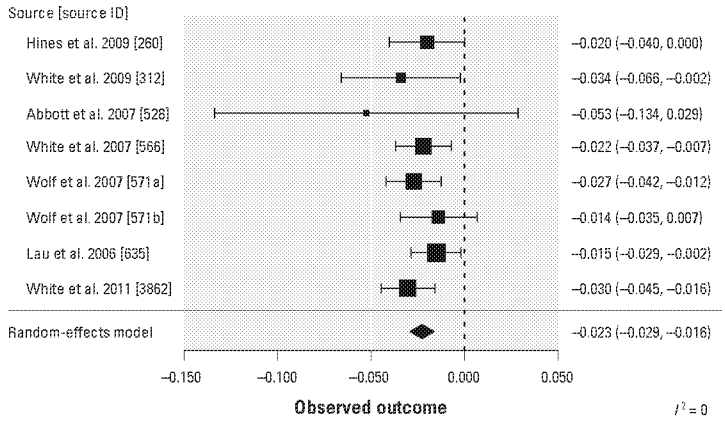

Figure 4. Meta-analysis results from a two-step mixed-effects model using combined relevant mouse studies in which dams were treated with PFOA via gavage and progeny weight was measured at or soon after birth. Data are presented as the mean (95% CI) change in body weight (g) per 1-unit increase in dose (mg/kg BW/day). Each box represents the dose–response slope estimate for a study; the midpoint of the box is the slope estimated for that study, and the box area is proportional to the weight given to each study in the meta-analysis. The diamond is centered at the overall meta-analysis slope estimate. Wolf (2007) was split into two data sets: *a*) cross-foster (exposure on GDs 1–17), and *b*) windows of sensitivity (exposure groups for GDs 7–17, GDs 10–17, GDs 13–17, and GDs 15–17).

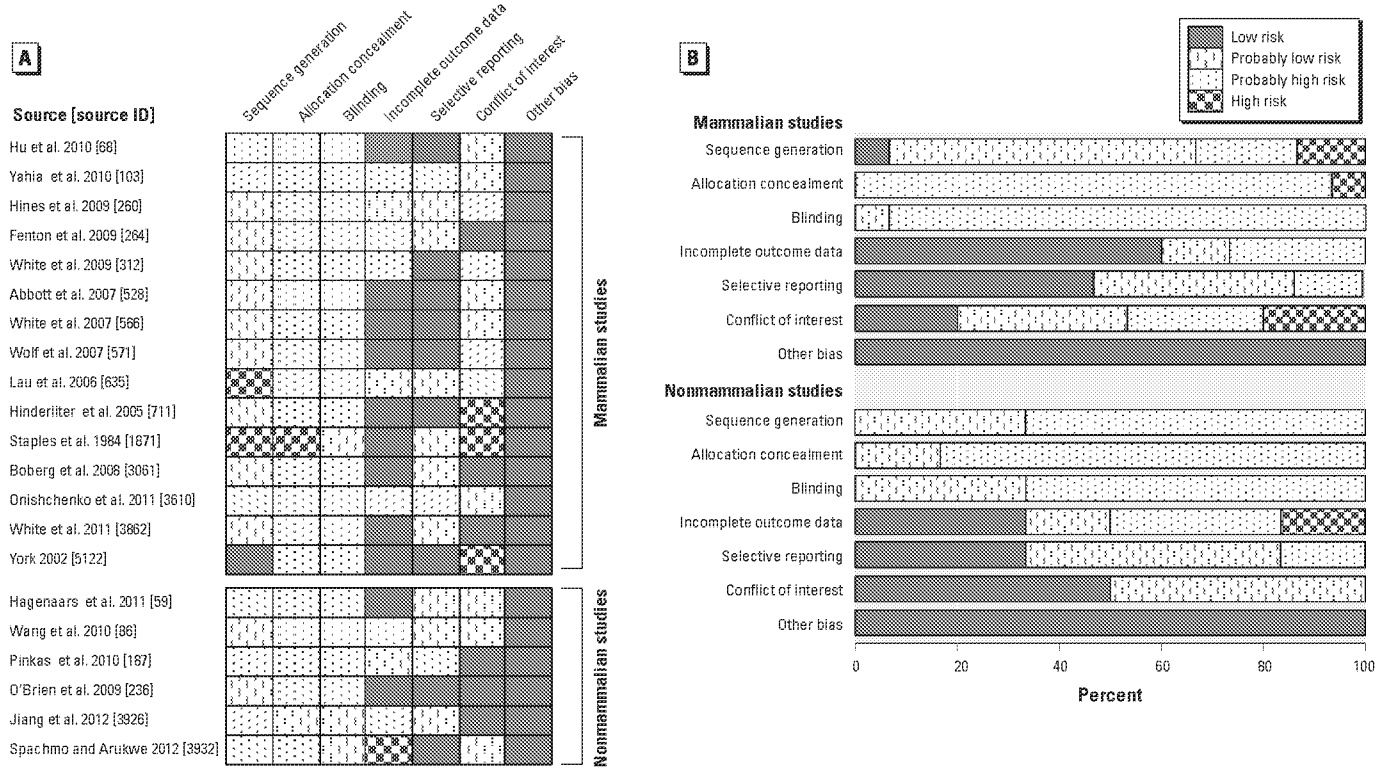

Figure 3. Summary of review authors' risk of bias judgments (low, probably low, probably high, and high risk) for each risk of bias item for each included study (*A*) and given as percentages across all included studies (*B*), separated into mammalian (n = 15) and nonmammalian (n = 6) studies.

shifting it to a smaller estimate of decreased birth weight. Although the data set of Abbott et al. (2007) had the largest effect estimate, removing that data set had little effect on the meta-analysis because of its small weight. The sensitivity analysis further demonstrated that the 95% CIs were also minimally affected and consistently did not include zero.

We created separate scatter plots to summarize all the mammalian study data for near-term, fetal weight measurements (Figure 5A) and for birth weight measurements (Figure 5B). The dose–response data for the nine studies not included in the meta-analysis showed mixed results, generally with lower doses showing increased weight compared with the control group (mostly nonsignificant) and higher doses showing decreased weight (some statistically significant and others not significant) (Figure 5B). The 95% CIs for the mean difference comparing birth weight in the treatment versus control group for each study are presented in Supplemental Material, Tables S46 and S47.

We also created scatter plots to summarize nonmammalian study data separately for weight measurements (Figure 6A) and for length measurements (Figure 6B). A qualitative evaluation of dose–response data showed mostly nonstatistically significant increases in body weight, even at the highest tested doses. The length data show mixed results, with two studies demonstrating statistically significant decreases in length and the other two studies showing nonsignificant increases in length. The 95% CIs for the mean difference comparing birth weight in the treatment versus control group for each study are presented in Supplemental Material, Tables S48 and S49.

Quality of evidence. We downgraded the overall quality rating of the mammalian evidence from "high" to "moderate" based on risk of bias across studies, because the majority of studies were deemed to have "probably high" risk of bias for allocation concealment and blinding. Our ratings and rationales for the overall quality of mammalian evidence are presented in Table 5.

We downgraded the overall quality rating of the nonmammalian evidence from "high" to "low" because of *a*) risk of bias across studies, given that most studies were deemed to have "probably high" risk of bias for the sequence generation, allocation concealment, and blinding domains; and *b*) indirectness—for the purposes of this case study, we did not have a rationale or evidence to support that all the nonmammalian species and their corresponding routes of exposure were directly applicable model systems for evaluating human fetal growth. Our ratings and rationales for the overall quality of nonmammalian evidence are presented in Table 6.

Strength of evidence rating. We excluded the nonmammalian data from the final strength of evidence rating. Our rationale was that the nonmammalian evidence was judged to be of low quality for the purposes of addressing our study question, and we had higher quality direct evidence on which to base a decision. Our strength of the evidence considerations were as follows:

- Quality of body of evidence: moderate
- Direction of effect estimates: decreasing birth weight with increasing exposure to PFOA
- Confidence in effect estimates: confidence based on the consistency of the results and overlapping CIs
- Other compelling attributes of the data that may influence certainty: none.

We compared these considerations with the definitions in Table 2 and concluded that the animal evidence is sufficient to conclude that exposure to PFOA or its salts adversely affect fetal growth in animals.

Discussion

Animal Evidence for PFOA and Fetal Growth

Based on this first application of the Navigation Guide systematic review methodology, we found "sufficient" evidence that fetal developmental exposure to PFOA or its salts reduces fetal growth in animals. Our finding that the data were "sufficient" was based on "moderate" quality mammalian evidence, reduction in mean offspring birth weight from dams exposed to increasing concentrations of PFOA during pregnancy, and our confidence in the effect based

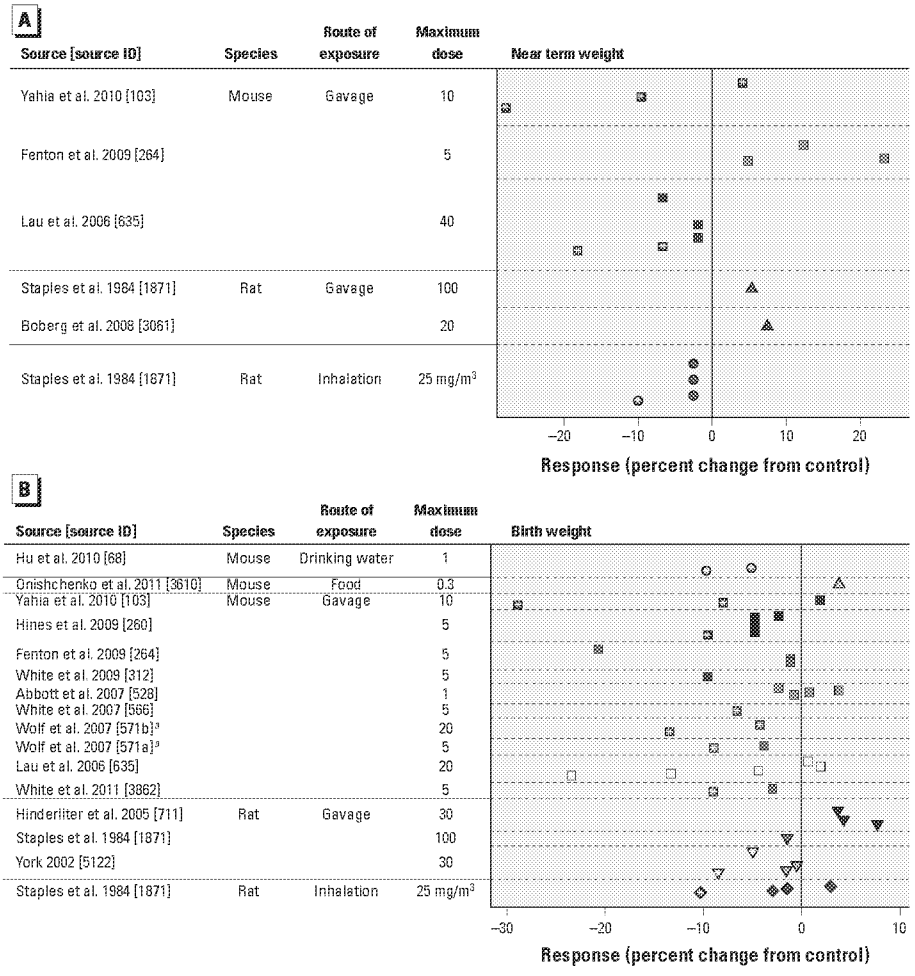

Figure 5. Combined scatter plots of response for each tested dose of PFOA for all included mammalian studies. Response was measured as the percentage of weight change for progeny (*A*) near-term or (*B*) at birth. Each color represents a different study (separated by dashed lines), and each symbol represents a different species or exposure route category. Multiple symbols of the same color represent responses at multiple tested doses within the same study. Doses are given in mg/kg BW/day unless otherwise specified. For each study, doses decrease as the *y*-axis increases and are scaled appropriately (i.e., larger vertical gaps indicate larger gaps between doses); the minimum dose for all studies is zero. See Supplemental Material, Tables S46 and S47, for the 95% CIs for the point estimates shown in the figure.

[a]Study split into two data sets: *a*) cross-foster (exposure on GDs 1–17), and *b*) windows of sensitivity (exposure groups for GDs 7–17, GDs 10–17, GDs 13–17, and GDs 15–17). *(Within symbols), $p < 0.05$ compared with the control group.

on the consistency of the results and overlapping CIs. Analysis of the scatter plots of the studies excluded from the meta-analysis supported that the majority of these studies also found consistently small reductions in measures of fetal growth following maternal exposure to PFOA.

From the meta-analysis of eight mouse gavage data sets, we estimated that exposure of pregnant mice to increasing concentrations of PFOA was associated with a change in mean pup birth weight of –0.023 g (95% CI: –0.029, –0.016) per 1-unit increase in dose (mg/kg BW/day). To assess the biological significance of this estimate, we pooled birth weight measurements from each of the eight control groups to estimate an overall mean birth weight of 1.57 g for the pups in control groups. A 0.023 g decrease in body weight is equivalent to an approximate 1.46% decrease in average body weight per 1-unit increase in PFOA dose. Thus, for example, according to this model, a dose of 10 mg/kg BW/day PFOA to pregnant dams is estimated to result in approximately a 15% decrease in the litter's average birth weight.

To address the heterogeneity of the available evidence, we limited the meta-analysis to data from mouse studies. The rationale for this decision was based in part on findings from pharmacokinetic studies documenting that the rate of elimination for PFOA is much faster for female rats compared with other mammalian species, including humans (Lau et al. 2007). Many of the studies included in our meta-analysis cited rate of elimination differences as a supporting reason for using mouse model systems. However, responses between mouse model systems may differ as well; evidence suggests that responses to PFOA may vary based on the mouse strain tested. One study noted that the 129S1/SvImJ strain was more sensitive to PFOA exposure compared with the CD-1 strain (Abbott et al. 2007). We included data from the 129S1/SvImJ strain in our meta-analysis because, in the absence of evidence supporting which mouse strain best matches human sensitivity to PFOA, there was no evidence to support a premise that humans are less sensitive than the most sensitive mouse. This is further supported by studies of agents known to cause reproductive toxicity, for which "humans appear to be as or more sensitive than the most sensitive animal species tested" (U.S. EPA 1996). Additionally, our sensitivity analysis found removing this study from the meta-analysis resulted in minimal changes in the meta-analysis estimate (< 2%) (data not shown).

The heterogeneity of the nonmammalian animal data precluded combining these studies quantitatively. Our identification of studies among such diverse species was unexpected, and for this case study, we combined all nonmammalian species into a single body of evidence. This did not impede decision making about toxicity of PFOA and fetal growth because more direct mammalian and human data were available. However, for other chemicals, heterogeneous indirect evidence may be the only data available on which to base a decision. This points to the need to anticipate and plan for the analysis of heterogeneous data—including whether it is appropriate to evaluate each species separately—and to determine relevance to human health beforehand in future protocols.

Application of the Navigation Guide Systematic Review Methodology

We found the application of the Navigation Guide method to be effective in producing a concise statement of health hazard in a systematic and transparent manner. Although our review did not identify any studies relevant to our study question that were published in languages other than English, it is difficult to predict in which cases excluding non-English studies may bias a systematic review (Sterne et al. 2011b). Therefore, for future reviews we would retain this strategy. Moreover, our systematic search identified > 1,900 studies that we did not find in a previous search that we conducted at the initiation of the project using traditional nonsystematic methods, and our improved search strategy nearly doubled the number of studies that met our prespecified inclusion criteria.

Despite a steep learning curve, designing and completing the search, eliminating duplicate records, screening studies, and extracting study characteristics and data took 2–3 months, including time to train review authors. Contact with study authors to obtain additional information took place over the

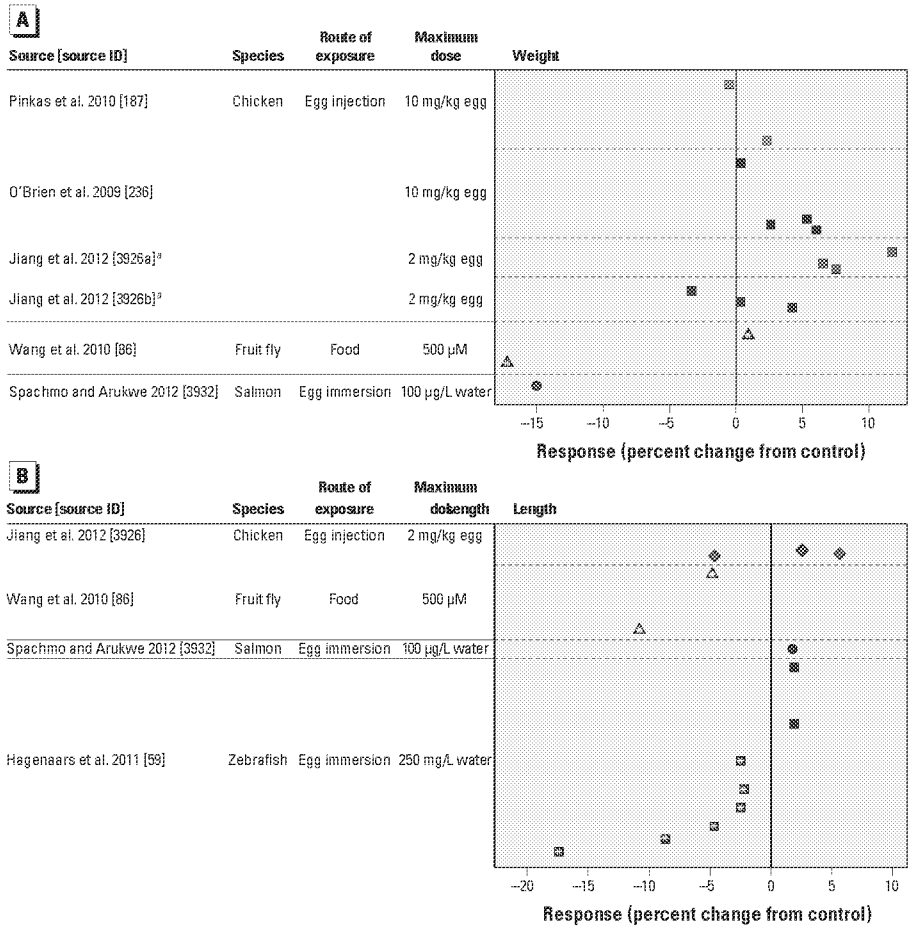

Figure 6. Combined scatter plots of response for each tested dose of PFOA for all included nonmammalian studies. Response was measured as the percentage of (*A*) weight change and (*B*) length change. Each color represents a different study (separated by dashed lines), and each symbol represents a different species or exposure route category. Multiple symbols of the same color represent responses at multiple tested doses within the same study. For each study, doses decrease as the *y*-axis increases and are scaled appropriately (i.e., larger vertical gaps indicate larger gaps between doses); the minimum dose for all studies is zero. See Supplemental Material, Tables S48 and S49, for the 95% CIs for the point estimates shown in the figure.

[a]Study split into two data sets based on time of outcome measurement: *a*) embryonic day 19, and *b*) 16–24 hr posthatching.

*(Within symbols), $p < 0.05$ compared with the control group.

course of approximately 3 months. Risk of bias assessment, data analysis, and evaluation of quality and strength of evidence took approximately 2–3 additional months.

An inevitable limitation of this first case study was that we were simultaneously developing and applying the method. As a result, we did not anticipate or define beforehand all the benchmarks we ultimately used for making judgments when rating the quality and strength of the evidence, and we found that our decision making was more difficult in the absence of prespecified definitions. To guide our judgments when assessing quality and strength of evidence factors that had not been prespecified, we *a*) sought an empirical basis for a judgment; *b*) conducted further analysis (i.e., sensitivity); *c*) relied on GRADE's principle to be conservative in the judgment of rating down; and *d*) always documented the rationale for our judgment. Anticipating and defining prespecified criteria for as many judgments as possible will improve the method; however, it seems unlikely that all judgments can be anticipated. Thus, the principles we used for judgments subsequent to the analysis will be integrated into future protocols to transparently allow for such circumstances.

Challenges in Translating Experimental Animal Evidence into Improved Health Outcomes

In applying the Navigation Guide systematic review methodology, we found that the high prevalence of suboptimal study design and reporting in experimental animal studies that has been empirically documented in the preclinical literature (Bebarta et al. 2003; Landis et al. 2012; Macleod et al. 2004, 2008; McPartland et al. 2007; van der Worp and Macleod 2011; van der Worp et al. 2007; Vesterinen et al. 2011) may also be prevalent in the experimental animal data that inform decision making in environmental health. In nearly all of the studies included in our review, direct evidence to support risk of bias ratings, such as clear descriptions of randomization or blinding methods, was missing. Furthermore, many studies failed to report some of the basic data necessary for interpretation of results and incorporation into meta-analysis. For example,

Table 5. Mammalian summary of findings, quality of evidence, and strength of evidence.

Factor	Rating	Basis
Risk of bias across studies	–1	"Allocation concealment" and "blinding" risks of bias were *a*) truly present, and *b*) these risks of bias are shown empirically to influence study outcome in preclinical experimental animal studies.
Indirectness	0	Mammalian data are empirically recognized as direct evidence of human health (Kimmel et al. 1984; U.S. EPA 1996) and there are no data to counteract this assumption.
Inconsistency	0	Point estimates across similar studies (e.g., mouse gavage) are consistent with overlapping confidence bounds. Estimates of change in birth weight from studies in the meta-analysis are consistently in the same direction and have low heterogeneity. Results are also consistent in magnitude and direction of effect estimates. Results of the meta-analysis do not appear to be strongly influenced by an individual study.
Imprecision	0	Mammalian data included in the meta-analysis showed relatively small CIs in final estimates. Although some studies did not report CIs, data show statistically significant responses at high doses—indicating small CIs.
Publication bias	0	We found no reason to suspect publication bias. The studies were consistent among their findings regardless of size and funding source; the search was comprehensive, and no unpublished studies were found that presented results out of the range of estimates reported by published studies.
Overall quality of evidence (initial rating is "high")	Moderate	"High" + (–1) = "moderate"
Summary of findings from meta-analysis	NA	Average change in birth weight = –0.023 g [–0.029, –0.016] per 1-unit increase in dose (mg/kg BW/day)
Summary of findings from qualitative analysis	NA	The dose–response data showed mixed results, generally with lower doses showing increased weight compared with the control group (mostly nonsignificant) and higher doses showing decreased weight (some statistically significant and other not significant).
Overall strength of evidence	Sufficient	

NA, not applicable. Ratings: –1, 1 level downgrade in quality. 0, no change in quality. Studies included in meta-analysis [source ID]: Abbott et al. 2007 [528], Hines et al. 2009 [260], Lau et al. 2006 [635] (birth weight data), White et al. 2007 [566], White et al. 2009 [312], White et al. 2011 [3862], and Wolf et al. 2007 [571] (cross-foster and windows of sensitivity data). Other studies [source ID]: Boberg 2008 et al. [3061], Fenton et al. 2009 [264], Hinderliter et al. 2005 [711], Hu et al. 2010 [68], Lau et al. 2006 [635] (fetal weight data), Onishchenko et al. 2011 [3610], Staples 1984 et al. [1871], Yahia et al. 2010 [103], and York 2002 [5122].

Table 6. Nonmammalian summary of findings, quality of evidence, and strength of evidence.

Factor	Rating	Basis
Risk of bias across studies	–1	"Sequence generation," "allocation concealment," and "blinding" risks of bias were: *a*) truly present; and *b*) these risks of bias are shown to empirically influence study outcome in preclinical experimental animal studies.
Indirectness	–1	We lacked an empirical basis supporting that these nonmammalian data were directly relevant to the human health outcome of interest, and the routes of exposure varied from how humans would be exposed to PFOA. Some evidence supports indirectness, in particular: Embryonic development in mammalian organisms (i.e., *in utero* development and live birth) is fundamentally different from development in nonmammalian organisms (i.e., development in egg and hatching), and the route of exposures for the nonmammalian organisms (i.e., eggs injected with or immersed in PFOA-containing solution) are not applicable to humans or other mammalian organisms.
Inconsistency	0	Results appear to divide based on measurement of outcome (weight vs. length); however, results are consistent between comparable studies (comparable for outcome, species, and exposure route).
Imprecision	0	The zebrafish and fruit fly data have a relatively large sample size, and while no confidence bounds are given, the effect estimates are reasonably close to each other (–5% to –20% change). Although some studies did not report CIs, data show statistically significant responses at high doses—indicating small CIs.
Publication bias	0	We found no reason to suspect publication bias. The search was comprehensive, the studies were of various sizes and had various funding sources, and no unpublished studies were found that presented results out of the range of estimates reported by published studies.
Overall quality of evidence (initial rating is "high")	Low	"High" + (–2) = "Low"
Summary of findings from qualitative analysis	NA	Dose–response data show mostly nonstatistically significant increases in body weight, even at the highest tested doses. The length data show mixed results, with two studies demonstrating statistically significant decreases in length and the other two studies showing statistically nonsignificant increases in length.

NA, not applicable. Ratings: –1, 1 level downgrade in quality. 0, no change in quality. Studies [source ID]: Hagenaars et al. 2011 [59], Jiang et al. 2012 [3926], O'Brien et al. 2009 [236], Pinkas et al. 2010 [187], Spachmo and Arukwe 2012 [3932], and Wang et al. 2010 [86].

multiple studies failed to report data such as the number of animals included in outcome measurements (e.g., number of litters assessed, number of pups per litter), details on how offspring were weighed (e.g., individually, as a whole litter), or the time point of outcome assessment (e.g., clear definition of PND1, monitoring of parturition). To create scatter plots and perform a meta-analysis, we needed to contact the lead author of every study to obtain missing data. Fortunately, authors for most of the studies responded, and many generously took the time and effort to provide raw data for inclusion in this review. Our follow-up with the authors indicated that many of these missing data were a result of deficiencies in reporting and point to the need to include contacting study authors as a step in the protocol.

These findings underscore the urgency of calls for improved experimental-animal study design and reporting in the preclinical arena (Beronius et al. 2014; Krauth et al. 2013; Landis et al. 2012; van der Worp and Macleod 2011; Vesterinen et al. 2010, 2011). To this end, a major stakeholder meeting by the National Institute of Neurological Disorders and Stroke found that at a minimum, studies should report on sample-size estimation, whether and how animals were randomized, whether investigators were blind to the treatment, and the handling of data (Landis et al. 2012). It will be important for environmental health scientists and journals that publish environmental health research to help support these nascent efforts to advance the translational relevance of animal evidence into improved health outcomes (Howells and Macleod 2013; Macleod et al. 2009; van der Worp et al. 2010; Vesterinen et al. 2011).

Summary and Conclusion

This case study documents that the Navigation Guide methodology can be used to effectively apply the rigor of evidence synthesis methods in use by the clinical sciences to questions in environmental health. The Navigation Guide methodology does not eliminate the need for expert judgment, but it does *a*) make clear the evidence that informs the authors' judgments and *b*) require transparency and an explicit accounting of the judgments involved.

In addition to this review of the animal evidence, a separate systematic review was conducted evaluating the human evidence relevant to PFOA exposure and fetal growth, which resulted in a "sufficient" evidence of toxicity rating (Johnson et al. 2014). In another paper, the strength of the evidence ratings from the nonhuman and human evidence were combined according to the factors specified in the Navigation Guide (Woodruff et al. 2011a), resulting in an overall conclusion by the review authors that human exposure to PFOA is "known to be toxic" to human reproduction and development based on "sufficient" evidence of decreased fetal growth in both human and nonhuman mammalian species (Lam et al. 2014). Together, these reviews demonstrate the utility of the Navigation Guide in systematically approaching a complex body of scientific evidence.

The ultimate goal of our efforts is to refine the Navigation Guide systematic review methodology across diverse streams of evidence and to support the development of recommendations for prevention in clinical and policy spheres. As has been demonstrated in the clinical field, the adoption of systematic and transparent methods to synthesize the scientific evidence in the environmental health field would speed incorporation of research into decision making.

References

Abbott BD, Wolf CJ, Schmid JE, Das KP, Zehr RD, Helfant L, et al. 2007. Perfluorooctanoic acid–induced developmental toxicity in the mouse is dependent on expression of peroxisome proliferator–activated receptor-alpha. Toxicol Sci 98:571–581.

Apelberg BJ, Goldman LR, Calafat AM, Herbstman JB, Kuklenyik Z, Heidler J, et al. 2007a. Determinants of fetal exposure to polyfluoroalkyl compounds in Baltimore, Maryland. Environ Sci Technol 41:3891–3897.

Apelberg BJ, Witter FR, Herbstman JB, Calafat AM, Halden RU, Needham LL, et al. 2007b. Cord serum concentrations of perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) in relation to weight and size at birth. Environ Health Perspect 115:1670–1676; doi:10.1289/ehp.10334.

ATSDR (Agency for Toxic Substances and Disease Registry). 2009. Draft Toxicological Profile for Perfluoroalkyls. Available: http://www.atsdr.cdc.gov/toxprofiles/tp200.pdf [accessed 13 February 2013].

Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. 2011. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 64:401–406.

Bebarta V, Luyten D, Heard K. 2003. Emergency medicine animal research: does use of randomization and blinding affect the results? Acad Emerg Med 10:684–687.

Beronius A, Molander L, Rudén C, Hanberg A. 2014. Facilitating the use of non-standard *in vivo* studies in health risk assessment of chemicals: a proposal to improve evaluation criteria and reporting. J Appl Toxicol 34:607–617; doi:10.1002/jat.2991.

Boberg J, Metzdorff S, Wortziger R, Axelstad M, Brokken L, Vinggaard AM, et al. 2008. Impact of diisobutyl phthalate and other PPAR agonists on steroidogenesis and plasma insulin and leptin levels in fetal rats. Toxicology 250:75–81.

Butenhoff JL, Kennedy GL Jr, Frame SR, O'Connor JC, York RG. 2004. The reproductive toxicology of ammonium perfluorooctanoate (APFO) in the rat. Toxicology 196:95–116.

Calafat AM, Kuklenyik Z, Reidy JA, Caudill SP, Tully JS, Needham LL. 2007. Serum concentrations of 11 polyfluoroalkyl compounds in the U.S. population: data from the National Health and Nutrition Examination Survey (NHANES). Environ Sci Technol 41:2237–2242.

CDC (Centers for Disease Control and Prevention). 2009. Fourth National Report on Human Exposure to Environmental Chemicals. Atlanta, GA:CDC. Available: http://www.cdc.gov/exposurereport/ [accessed 6 February 2013].

Deeks JJ, Higgins JPT, Altman DG, eds. 2011. Chapter 9: Analysing data and undertaking meta-analyses. In: Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (Higgins JPT, Green S, eds). Available: http://www.cochrane-handbook.org [accessed 15 April 2013].

Fei C, McLaughlin JK, Tarone RE, Olsen J. 2007. Perfluorinated chemicals and fetal growth: a study within the Danish National Birth Cohort. Environ Health Perspect 115:1677–1682; doi:10.1289/ehp.10506.

Fei C, McLaughlin JK, Tarone RE, Olsen J. 2008. Fetal growth indicators and perfluorinated chemicals: a study in the Danish National Birth Cohort. Am J Epidemiol 168:66–72.

Fenton SE, Reiner JL, Nakayama SF, Delinsky AD, Stanko JP, Hines EP, et al. 2009. Analysis of PFOA in dosed CD-1 mice. Part 2. Disposition of PFOA in tissues and fluids from pregnant and lactating mice and their pups. Reprod Toxicol 27:365–372.

Gluckman PD, Hanson MA. 2006. Developmental Origins of Health and Disease. Cambridge, UK:Cambridge University Press.

Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. 2011. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol 64:407–415.

Hagenaars A, Vergauwen L, De Coen W, Knapen D. 2011. Structure-activity relationship assessment of four perfluorinated chemicals using a prolonged zebrafish early life stage test. Chemosphere 82:764–772.

Hamm MP, Cherry NM, Chan E, Martin JW, Burstyn I. 2010. Maternal exposure to perfluorinated acids and fetal growth. J Expo Sci Environ Epidemiol 20:589–597.

Higgins JPT, Altman DJ, Sterne JAC, eds. 2011. Chapter 8: Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (Higgins JPT, Green S, eds). Available: http://www.cochrane-handbook.org [accessed 7 February 2013].

Higgins JPT, Deeks JJ, eds. 2011. Chapter 7: Selecting studies and collecting data. In: Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (Higgins JPT, Green S, eds). Available: http://www.cochrane-handbook.org [accessed 7 February 2013].

Higgins JPT, Thompson SG, Deeks JJ, Altman DG. 2003. Measuring inconsistency in meta-analyses. BMJ 327:557–560.

Hinderliter PM, Mylchreest E, Gannon SA, Butenhoff JL, Kennedy GL Jr. 2005. Perfluorooctanoate: placental and lactational transport pharmacokinetics in rats. Toxicology 211:139–148.

Hines EP, White SS, Stanko JP, Gibbs-Flournoy EA, Lau C, Fenton SE. 2009. Phenotypic dichotomy following developmental exposure to perfluorooctanoic acid (PFOA) in female CD-1 mice: low doses induce elevated serum leptin and insulin, and overweight in mid-life. Mol Cell Endocrinol 304:97–105.

Hooijmans CR, Leenaars M, Ritskes-Hoitinga M. 2010a. A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. Altern Lab Anim 38:167–182.

Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. 2010b. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. Lab Anim 44:170–175.

Howells DW, Macleod MR. 2013. Evidence-based translational medicine. Stroke 44:1466–1471.

Hu Q, Strynar MJ, DeWitt JC. 2010. Are developmentally exposed C57BL/6 mice insensitive to suppression of TDAR by PFOA? J Immunotoxicol 7:344–349.

IARC (International Agency for Research on Cancer). 2006. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Preamble. Available: http://monographs.iarc.fr/ENG/Preamble/CurrentPreamble.pdf [accessed 7 February 2013].

Institute of Medicine. 2007. Preterm Birth: Causes, Consequences, and Prevention. Washington, DC:National Academies Press. Available: http://books.nap.edu/openbook.php?record_id=11622 [accessed 25 August 2014].

Jiang Q, Lust RM, Strynar MJ, Dagnino S, DeWitt JC. 2012. Perfluorooctanoic acid induces developmental cardiotoxicity in chicken embryos and hatchlings. Toxicology 293:97–106.

Johnson PI, Sutton P, Atchley D, Koustas E, Lam J, Sen S, et al. 2014. The Navigation Guide—evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. Environ Health Perspect 122:1028–1039; doi:10.1289/ehp.1307893.

Kannan K, Corsolini S, Falandysz J, Fillmann G, Kumar KS, Loganathan BG, et al. 2004. Perfluorooctanesulfonate and related fluorochemicals in human blood from several countries. Environ Sci Technol 38:4489–4495.

Kato K, Wong LY, Jia LT, Kuklenyik Z, Calafat AM. 2011. Trends

in exposure to polyfluoroalkyl chemicals in the U.S. population: 1999–2008. Environ Sci Technol 45:8037–8045.

Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG. 2010. Animal research: reporting *in vivo* experiments: the ARRIVE guidelines. Br J Pharmacol 160:1577–1579.

Kimmel CA, Holson JF, Hogue CJ, Carlo GL. 1984. Reliability of Experimental Studies for Predicting Hazards to Human Development. NCTR Technical Report for Experiment 6015. Jefferson, AR:National Center for Toxicological Research.

Koustas E, Lam J, Sutton P, Johnson P, Atchley D, Sen S, et al. 2013. Applying the Navigation Guide: Case Study #1. The Impact of Developmental Exposure to Perfluorooctanoic Acid (PFOA) On Fetal Growth. A Systematic Review of the Non-Human Evidence: Final Protocol. Available: http://prhe.ucsf.edu/prhe/pdfs/PFOA%20NON-HUMAN%20PROTOCOL.pdf [accessed 26 August 2014].

Krauth D, Woodruff TJ, Bero L. 2013. Instruments for Assessing Risk of Bias and Other Methodological Criteria of Published Animal Studies: A Systematic Review. Environ Health Perspect 121:985–992; doi:10.1289/ehp.1206389.

Lam J, Koustas E, Sutton P, Johnson PI, Atchley DS, Sen S, et al. 2014. The Navigation Guide----evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth. Environ Health Perspect 122:1040–1051; doi:10.1289/ehp.1307923.

Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, et al. 2012. A call for transparent reporting to optimize the predictive value of preclinical research. Nature 490:187–191.

Lau C, Anitole K, Hodes C, Lai D, Pfahles-Hutchens A, Seed J. 2007. Perfluoroalkyl acids: a review of monitoring and toxicological findings. Toxicol Sci 99:366–394.

Lau C, Butenhoff JL, Rogers JM. 2004. The developmental toxicity of perfluoroalkyl acids and their derivatives. Toxicol Appl Pharmacol 198:231–241.

Lau C, Thibodeaux JR, Hanson RG, Narotsky MG, Rogers JM, Lindstrom AB, et al. 2006. Effects of perfluorooctanoic acid exposure during pregnancy in the mouse. Toxicol Sci 90:510–518.

Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, Bath PM, et al. 2009. Good laboratory practice: preventing introduction of bias at the bench. Stroke 40:e50–e52; doi:10.1161/STROKEAHA.108.525386.

Macleod MR, O'Collins T, Howells DW, Donnan GA. 2004. Pooling of animal experimental data reveals influence of study design and publication bias. Stroke 35:1203–1208.

Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA. 2008. Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. Stroke 39:2824–2829.

McPartland JM, Glass M, Pertwee RG. 2007. Meta-analysis of cannabinoid ligand binding affinity and receptor distribution: interspecies differences. Br J Pharmacol 152:583–593.

Midasch O, Drexler H, Hart N, Beckmann MW, Angerer J. 2007. Transplacental exposure of neonates to perfluorooctanesulfonate and perfluorooctanoate: a pilot study. Int Arch Occup Environ Health 80:643–648.

Mondal D, Lopez-Espinosa MJ, Armstrong B, Stein CR, Fletcher T. 2012. Relationships of perfluorooctanoate and perfluorooctane sulfonate serum concentrations between mother–child pairs in a population with perfluorooctanoate exposure from drinking water. Environ Health Perspect 120:752–757; doi:10.1289/ehp.1104538.

Monroy R, Morrison K, Teo K, Atkinson S, Kubwabo C, Stewart B, et al. 2008. Serum levels of perfluoroalkyl compounds in human maternal and umbilical cord blood samples. Environ Res 108:56–62.

Mylchreest E. 2003. PFOA: Lactational and Placental Transport Pharmacokinetic Study in Rats. Laboratory Project ID: DuPont-13309. U.S. EPA Administrative Record 226–1551. Newark, DE:Haskell Laboratory for Health and Environmental Sciences. Available: http://www.epa.gov/oppt/tsca8e/pubs/8ehq/2004/jan04/8ehq_0104_00373ay.pdf [accessed 10 September 2014].

O'Brien JM, Crump D, Mundy LJ, Chu S, McLaren KK, Vongphachan V, et al. 2009. Pipping success and liver mRNA expression in chicken embryos exposed *in ovo* to C_8 and C_{11} perfluorinated carboxylic acids and C_{10} perfluorinated sulfonate. Toxicol Lett 190:134–139.

O'Connor D, Green S, Higgins JPT, eds. 2011. Chapter 5: Defining the review question and developing criteria for including studies. In: Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (Higgins JPT, Green S, eds). Available: http://www.cochrane-handbook.org [accessed 15 April 2013].

Olsen GW, Burris JM, Ehresman DJ, Froehlich JW, Seacat AM, Butenhoff JL, et al. 2007. Half-life of serum elimination of perfluorooctanesulfonate, perfluorohexanesulfonate, and perfluorooctanoate in retired fluorochemical production workers. Environ Health Perspect 115:1298–1305; doi:10.1289/ehp.10009.

Onishchenko N, Fischer C, Wan Ibrahim WN, Negri S, Spulber S, Cottica D, et al. 2011. Prenatal exposure to PFOS or PFOA alters motor function in mice in a sex-related manner. Neurotox Res 19:452–461.

Pinkas A, Slotkin TA, Brick-Turin Y, Van der Zee EA, Yanai J. 2010. Neurobehavioral teratogenicity of perfluorinated alkyls in an avian model. Neurotoxicol Teratol 32:182–186.

Roseman M, Milette K, Bero LA, Coyne JC, Lexchin J, Turner EH, et al. 2011. Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments. JAMA 305:1008–1017.

Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D. 2007. Update on the methods of the U.S. Preventative Services Task Force: estimating certainty and magnitude of net benefit. Ann Intern Med 147:871–875.

Spachmo B, Arukwe A. 2012. Endocrine and developmental effects in Atlantic salmon (*Salmo salar*) exposed to perfluorooctane sulfonic or perfluorooctane carboxylic acids. Aquat Toxicol 108:112–124.

Staples RE, Burgess BA, Kerns WD. 1984. The embryo-fetal toxicity and teratogenic potential of ammonium perfluorooctanoate (APFO) in the rat. Fundam Appl Toxicol 4:429–440.

Sterne JAC, Egger M, Moher D, eds. 2011b. Chapter 10: Addressing reporting biases. In: Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (Higgins JPT, Green S, eds). Available: http://www.cochrane-handbook.org [accessed 22 April 2013].

Sterne JAC, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. 2011a. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 343:d4002; doi:10.1136/bmj.d4002.

U.S. EPA (U.S. Environmental Protection Agency). 1991. Guidelines for Developmental Toxicity Risk Assessment. Available: http://www.epa.gov/raf/publications/pdfs/DEVTOX.PDF [accessed 10 September 2014].

U.S. EPA (U.S. Environmental Protection Agency). 1996. Guidelines for Reproductive Toxicity Risk Assessment. Available: http://www.epa.gov/raf/publications/pdfs/REPRO51.PDF [accessed 10 September 2014].

U.S. EPA (U.S. Environmental Protection Agency). 2008. EPA Announces Substantial Decrease of PFOA [News Release]. Available: http://yosemite.epa.gov/opa/admpress.nsf/68b5f2d54f3eefd28525701500517fbf/8f9dbdd044050f71852573e50064439f!OpenDocument [accessed 3 February 2013].

U.S. EPA (U.S. Environmental Protection Agency). 2009. Long-Chain Perfluorinated Chemicals (PFCs) Action Plan. Available: http://www.epa.gov/oppt/existingchemicals/pubs/pfcs_action_plan1230_09.pdf [accessed 3 February 2013].

U.S. EPA (U.S. Environmental Protection Agency). 2012. Industry Progressing in Voluntary Effort to Reduce Toxic Chemicals / Companies Have Successfully Developed over 150 Alternatives to Long-Chain Perfluorinated Chemicals [News Release:]. Available: http://yosemite.epa.gov/opa/admpress.nsf/79c090e81f05787385257 81f0043619b/335d29e8d5e40864852579a0005f3760!OpenDocument [accessed 11 February 2013].

U.S. EPA (U.S. Environmental Protection Agency). 2013a. 2012 Annual Progress Reports. Available: http://www.epa.gov/oppt/pfoa/pubs/stewardship/preports6.html [accessed 25 September 2013].

U.S. EPA (U.S. Environmental Protection Agency). 2013b. America's Children and the Environment. 3rd ed. Available: http://www.epa.gov/ace/publications/ACE3_2013.pdf [accessed 19 February 2013].

U.S. EPA (U.S. Environmental Protection Agency). 2014. Emerging Contaminants–Perfluorooctane Sulfonate (PFOS) and Perfluorooctanoic Acid (PFOA). Available: http://www2.epa.gov/sites/production/files/2014-04/documents/factsheet_contaminant_pfos_pfoa_march2014.pdf [accessed 10 September 2014].

van der Worp HB, Howells DW, Sena ES, Porritt MJ, Rewell S, O'Collins V, et al. 2010. Can animal models of disease reliably inform human studies? PLoS Med 7:e1000245; doi:10.1371/journal.pmed.1000245.

van der Worp HB, Macleod MR. 2011. Preclinical studies of human disease: time to take methodological quality seriously. J Mol Cell Cardiol 51:449–450.

van der Worp HB, Sena ES, Donnan GA, Howells DW, Macleod MR. 2007. Hypothermia in animal models of acute ischaemic stroke: a systematic review and meta-analysis. Brain 130:3063–3074.

Vesterinen HM, Egan K, Deister A, Schlattmann P, Macleod MR, Dirnagl U. 2011. Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the *Journal of Cerebral Blood Flow and Metabolism*. J Cereb Blood Flow Metab 31:1064–1072.

Vesterinen HM, Sena ES, ffrench-Constant C, Williams A, Chandran S, Macleod MR. 2010. Improving the translational hit of experimental treatments in multiple sclerosis. Mult Scler 16:1044–1055.

Viechtbauer W. 2010. Conducting meta-analyses in R with the metafor package. J Stat Softw 36:1–48.

Wang J, Li Y, Liu Y, Zhang H, Dai J. 2010. Disturbance of perfluorooctanoic acid on development and behavior in Drosophila larvae. Environ Toxicol Chem 29:2117–2122.

Washino N, Saijo Y, Sasaki S, Kato S, Ban S, Konishi K, et al. 2009. Correlations between prenatal exposure to perfluorinated chemicals and reduced fetal growth. Environ Health Perspect 117:660–667; doi:10.1289/ehp.11681.

White SS, Calafat AM, Kuklenyik Z, Villanueva L, Zehr RD, Helfant L, et al. 2007. Gestational PFOA exposure of mice is associated with altered mammary gland development in dams and female offspring. Toxicol Sci 96:133–144.

White SS, Kato K, Jia LT, Basden BJ, Calafat AM, Hines EP, et al. 2009. Effects of perfluorooctanoic acid on mouse mammary gland development and differentiation resulting from cross-foster and restricted gestational exposures. Reprod Toxicol 27:289–298.

White SS, Stanko JP, Kato K, Calafat AM, Hines EP, Fenton SE. 2011. Gestational and chronic low-dose PFOA exposures and mammary gland growth and differentiation in three generations of CD-1 mice. Environ Health Perspect 119:1070–1076; doi:10.1289/ehp.1002741.

Wolf CJ, Fenton SE, Schmid JE, Calafat AM, Kuklenyik Z, Bryant XA, et al. 2007. Developmental toxicity of perfluorooctanoic acid in the CD-1 mouse after cross-foster and restricted gestational exposures. Toxicol Sci 95:462–473.

Woodruff TJ, Sutton P. 2014. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect 122:1007–1014; doi:10.1289/ehp.1307175.

Woodruff TJ, Sutton P, Navigation Guide Work Group. 2011a. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Aff (Millwood) 30:931–937.

Woodruff TJ, Zota AR, Schwartz JM. 2011b. Environmental chemicals in pregnant women in the United States: NHANES 2003–2004. Environ Health Perspect 119:878–885; doi:10.1289/ehp.1002727.

Yahia D, El-Nasser MA, Abedel-Latif M, Tsukuba C, Yoshida M, Sato I, et al. 2010. Effects of perfluorooctanoic acid (PFOA) exposure to pregnant mice on reproduction. J Toxicol Sci 35:527–533.

York RG. 2002. Oral (Gavage) Two-Generation (One Litter per Generation) Reproduction Study of Ammonium Perfluorooctanoate (APFO) in Rats. Sponsor's Study No. T-6889.6. U.S. EPA Administrative Record 226–1092. Horsham, PA:Argus Research.

York RG, Kennedy GL Jr, Olsen GW, Butenhoff JL. 2010. Male reproductive system parameters in a two-generation reproduction study of ammonium perfluorooctanoate in rats and human relevance. Toxicology 271:64–72.

Environmental Health Perspectives doi:10.1289/ehp.1307177

Supplemental Material

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Nonhuman Evidence for PFOA Effects on Fetal Growth

Erica Koustas, Juleen Lam, Patrice Sutton, Paula I. Johnson, Dylan S. Atchley, Saunak Sen, Karen A. Robinson, Daniel A. Axelrad, and Tracey J. Woodruff

Table of Contents **Page**

Table S1: PubMed search terms.

Search	PubMed
#1 Substance terms	335-67-1 [rn] OR perfluorooctanoic acid [nm] OR (perfluorooctanoic acid [tiab] OR perfluorooctanoic acids [tiab]) OR (perfluoroctanoic acid [tiab] OR perfluoroctanoic acids [tiab]) OR (perfluoro-n-octanoic acid [tiab] OR perfluoro-n-octanoic acids [tiab]) OR (pentadecafluorooctanoic acid [tiab] OR pentadecafluorooctanoic acids [tiab]) OR APFO [tiab] OR (perfluorinated [tiab] AND octanoic acid [tiab]) OR (perfluorinated [tiab] AND octanoic acids [tiab]) OR (perfluorooctanoate [tiab] OR perfluorooctanoates [tiab]) OR perfluorooctanoyl chloride [tiab] OR PFOA [tiab] OR (fluorinated telomer alcohol [tiab] OR fluorinated telomer alcohols [tiab]) OR (fluoro-telomer alcohol [tiab] OR fluoro-telomer alcohols [tiab]) OR (fluorocarbon emulsion [tiab] OR fluorocarbon emulsions [tiab]) OR (perfluorocarbon [tiab] OR perfluorocarbons [tiab]) OR (fluorocarbon polymer [tiab] OR fluorocarbon polymers [tiab]) OR (fluorinated polymer [tiab] OR fluorinated polymers [tiab]) OR octanoic acids [mh] OR (octanoic acid [tiab] OR octanoic acids [tiab]) OR caprylates [mh] OR (caprylate [tiab] OR caprylates [tiab]) OR (polyfluoroalkyl [tiab] OR polyfluoroalkyls [tiab] OR polyfluoroalkylated [tiab]) OR PFAA [tiab] OR (perfluoroalkyl chemical [tiab] OR perfluoroalkyl chemicals [tiab]) OR (c8 [tiab] AND perfluorinated [tiab]) OR (fluoropolymer [tiab] OR fluoropolymers [tiab] OR fluoropolymeric [tiab]) OR (fluorosurfactant [tiab] OR fluorosurfactants [tiab]) OR (perfluorochemical [tiab] OR perfluorochemicals [tiab]) OR PFCs [tiab] OR (perfluoroalkyl carboxylate [tiab] OR perfluoroalkyl carboxylates [tiab]) OR (perfluorocarboxylate [tiab] OR perfluorocarboxylates [tiab]) OR PFCA [tiab] OR (perfluorinated carboxylic acid [tiab] OR perfluorinated carboxylic acids [tiab]) OR FC 143 [tiab] OR (pentadecafluorooctanoate [tiab] OR pentadecafluorooctanoates [tiab])
#2 Experimental animal terms (modified from Hooijmans et al. 2010)	("animal experimentation"[MeSH Terms] OR "models, animal"[MeSH Terms] OR "invertebrates"[MeSH Terms] OR "Animals"[Mesh:noexp] OR "animal population groups"[MeSH Terms] OR "chordata"[MeSH Terms:noexp] OR "chordata, nonvertebrate"[MeSH Terms] OR "vertebrates"[MeSH Terms:noexp] OR "amphibians"[MeSH Terms] OR "birds"[MeSH Terms] OR "fishes"[MeSH Terms] OR "reptiles"[MeSH Terms] OR "mammals"[MeSH Terms:noexp] OR "primates"[MeSH Terms:noexp] OR "artiodactyla"[MeSH Terms] OR "carnivora"[MeSH Terms] OR "cetacea"[MeSH Terms] OR "chiroptera"[MeSH Terms] OR "elephants"[MeSH Terms] OR "hyraxes"[MeSH Terms] OR "insectivora"[MeSH Terms] OR "lagomorpha"[MeSH Terms] OR "marsupialia"[MeSH Terms] OR "monotremata"[MeSH Terms] OR "perissodactyla"[MeSH Terms] OR "rodentia"[MeSH Terms] OR "scandentia"[MeSH Terms] OR "sirenia"[MeSH Terms] OR "xenarthra"[MeSH Terms] OR "haplorhini"[MeSH Terms:noexp] OR "strepsirhini"[MeSH Terms] OR "platyrrhini"[MeSH Terms] OR "tarsii"[MeSH Terms] OR "catarrhini"[MeSH Terms:noexp] OR "cercopithecidae"[MeSH Terms] OR "hylobatidae"[MeSH Terms] OR "hominidae"[MeSH Terms:noexp] OR "gorilla gorilla"[MeSH Terms] OR "pan paniscus"[MeSH Terms] OR "pan troglodytes"[MeSH Terms] OR "pongo pygmaeus"[MeSH Terms]) OR (animals[tiab] OR animal[tiab] OR mice[Tiab] OR mus[Tiab] OR mouse[Tiab] OR murine[Tiab] OR woodmouse[tiab] OR rats[Tiab] OR rat[Tiab] OR murinae[Tiab] OR muridae[Tiab] OR cottonrat[tiab] OR cottonrats[tiab] OR hamster[tiab] OR hamsters[tiab] OR cricetinae[tiab] OR rodentia[Tiab] OR rodent[Tiab] OR rodents[Tiab] OR pigs[Tiab] OR pig[Tiab] OR swine[tiab] OR swines[tiab] OR piglets[tiab] OR piglet[tiab] OR boar[tiab] OR boars[tiab] OR "sus scrofa"[tiab] OR ferrets[tiab] OR ferret[tiab] OR polecat[tiab] OR polecats[tiab] OR "mustela putorius"[tiab] OR "guinea pigs"[Tiab] OR "guinea pig"[Tiab] OR cavia[Tiab] OR callithrix[Tiab] OR marmoset[Tiab] OR marmosets[Tiab] OR cebuella[Tiab] OR hapale[Tiab] OR octodon[Tiab] OR chinchilla[Tiab] OR chinchillas[Tiab] OR gerbillinae[Tiab] OR gerbil[Tiab] OR gerbils[Tiab] OR jird[Tiab] OR jirds[Tiab] OR merione[Tiab] OR meriones[Tiab] OR

Search	PubMed
	rabbits[Tiab] OR rabbit[Tiab] OR hares[Tiab] OR hare[Tiab] OR diptera[Tiab] OR flies[Tiab] OR fly[Tiab] OR dipteral[Tiab] OR drosphila[Tiab] OR drosophilidae[Tiab] OR cats[Tiab] OR cat[Tiab] OR carus[Tiab] OR felis[Tiab] OR nematoda[Tiab] OR nematode[Tiab] OR nematodes[Tiab] OR sipunculida[Tiab] OR dogs[Tiab] OR dog[Tiab] OR canine[Tiab] OR canines[Tiab] OR canis[Tiab] OR sheep[Tiab] OR sheeps[Tiab] OR mouflon[Tiab] OR mouflons[Tiab] OR ovis[Tiab] OR goats[Tiab] OR goat[Tiab] OR capra[Tiab] OR capras[Tiab] OR rupicapra[Tiab] OR chamois[Tiab] OR haplorhini[Tiab] OR monkey[Tiab] OR monkeys[Tiab] OR anthropoidea[Tiab] OR anthropoids[Tiab] OR saguinus[Tiab] OR tamarin[Tiab] OR tamarins[Tiab] OR leontopithecus[Tiab] OR hominidae[Tiab] OR ape[Tiab] OR apes[Tiab] OR pan[Tiab] OR paniscus[Tiab] OR "pan paniscus"[Tiab] OR bonobo[Tiab] OR bonobos[Tiab] OR troglodytes[Tiab] OR "pan troglodytes"[Tiab] OR gibbon[Tiab] OR gibbons[Tiab] OR siamang[Tiab] OR siamangs[Tiab] OR nomascus[Tiab] OR symphalangus[Tiab] OR chimpanzee[Tiab] OR chimpanzees[Tiab] OR prosimians[Tiab] OR "bush baby"[Tiab] OR prosimian[Tiab] OR bush babies[Tiab] OR galagos[Tiab] OR galago[Tiab] OR pongidae[Tiab] OR gorilla[Tiab] OR gorillas[Tiab] OR pongo[Tiab] OR pygmaeus[Tiab] OR "pongo pygmaeus"[Tiab] OR orangutans[Tiab] OR lemur[Tiab] OR lemurs[Tiab] OR lemuridae[Tiab] OR horse[Tiab] OR horses[Tiab] OR equus[Tiab] OR cow[Tiab] OR calf[Tiab] OR bull[Tiab] OR chicken[Tiab] OR chickens[Tiab] OR gallus[Tiab] OR quail[Tiab] OR bird[Tiab] OR birds[Tiab] OR quails[Tiab] OR poultry[Tiab] OR poultries[Tiab] OR fowl[Tiab] OR fowls[Tiab] OR reptile[Tiab] OR reptilia[Tiab] OR reptiles[Tiab] OR snakes[Tiab] OR snake[Tiab] OR lizard[Tiab] OR lizards[Tiab] OR alligator[Tiab] OR alligators[Tiab] OR crocodile[Tiab] OR crocodiles[Tiab] OR turtle[Tiab] OR turtles[Tiab] OR amphibian[Tiab] OR amphibians[Tiab] OR amphibia[Tiab] OR frog[Tiab] OR frogs[Tiab] OR bombina[Tiab] OR salientia[Tiab] OR toad[Tiab] OR toads[Tiab] OR "epidalea calamita"[Tiab] OR salamander[Tiab] OR salamanders[Tiab] OR eel[Tiab] OR eels[Tiab] OR fish[Tiab] OR fishes[Tiab] OR pisces[Tiab] OR catfish[Tiab] OR catfishes[Tiab] OR siluriformes[Tiab] OR arius[Tiab] OR heteropneustes[Tiab] OR sheatfish[Tiab] OR perch[Tiab] OR perches[Tiab] OR percidae[Tiab] OR perca[Tiab] OR trout[Tiab] OR trouts[Tiab] OR char[Tiab] OR chars[Tiab] OR salvelinus[Tiab] OR "fathead minnow"[Tiab] OR minnow[Tiab] OR cyprinidae[Tiab] OR carps[Tiab] OR carp[Tiab] OR zebrafish[Tiab] OR zebrafishes[Tiab] OR goldfish[Tiab] OR goldfishes[Tiab] OR guppy[Tiab] OR guppies[Tiab] OR chub[Tiab] OR chubs[Tiab] OR tinca[Tiab] OR barbels[Tiab] OR barbus[Tiab] OR pimephales[Tiab] OR promelas[Tiab] OR "poecilia reticulata"[Tiab] OR mullet[Tiab] OR mullets[Tiab] OR seahorse[Tiab] OR seahorses[Tiab] OR mugil curema[Tiab] OR atlantic cod[Tiab] OR shark[Tiab] OR sharks[Tiab] OR catshark[Tiab] OR anguilla[Tiab] OR salmonid[Tiab] OR salmonids[Tiab] OR whitefish[Tiab] OR whitefishes[Tiab] OR salmon[Tiab] OR salmons[Tiab] OR sole[Tiab] OR solea[Tiab] OR "sea lamprey"[Tiab] OR lamprey[Tiab] OR lampreys[Tiab] OR pumpkinseed[Tiab] OR sunfish[Tiab] OR sunfishes[Tiab] OR tilapia[Tiab] OR tilapias[Tiab] OR turbot[Tiab] OR turbots[Tiab] OR flatfish[Tiab] OR flatfishes[Tiab] OR sciuridae[Tiab] OR squirrel[Tiab] OR squirrels[Tiab] OR chipmunk[Tiab] OR chipmunks[Tiab] OR suslik[Tiab] OR susliks[Tiab] OR vole[Tiab] OR voles[Tiab] OR lemming[Tiab] OR lemmings[Tiab] OR muskrat[Tiab] OR muskrats[Tiab] OR lemmus[Tiab] OR otter[Tiab] OR otters[Tiab] OR marten[Tiab] OR martens[Tiab] OR martes[Tiab] OR weasel[Tiab] OR badger[Tiab] OR badgers[Tiab] OR ermine[Tiab] OR mink[Tiab] OR minks[Tiab] OR sable[Tiab] OR sables[Tiab] OR gulo[Tiab] OR gulos[Tiab] OR wolverine[Tiab] OR wolverines[Tiab] OR mustela[Tiab] OR llama[Tiab] OR llamas[Tiab] OR alpaca[Tiab] OR alpacas[Tiab] OR camelid[Tiab] OR camelids[Tiab] OR guanaco[Tiab] OR guanacos[Tiab] OR chiroptera[Tiab] OR chiropteras[Tiab] OR bat[Tiab] OR bats[Tiab] OR fox[Tiab] OR foxes[Tiab] OR iguana[Tiab] OR iguanas[Tiab] OR xenopus laevis[Tiab] OR parakeet[Tiab] OR

Search	PubMed
	parakeets[Tiab] OR parrot[Tiab] OR parrots[Tiab] OR donkey[Tiab] OR donkeys[Tiab] OR mule[Tiab] OR mules[Tiab] OR zebra[Tiab] OR zebras[Tiab] OR shrew[Tiab] OR shrews[Tiab] OR bison[Tiab] OR bisons[Tiab] OR buffalo[Tiab] OR buffaloes[Tiab] OR deer[Tiab] OR deers[Tiab] OR bear[Tiab] OR bears[Tiab] OR panda[Tiab] OR pandas[Tiab] OR "wild hog"[Tiab] OR "wild boar"[Tiab] OR fitchew[Tiab] OR fitch[Tiab] OR beaver[Tiab] OR beavers[Tiab] OR jerboa[Tiab] OR jerboas[Tiab] OR capybara[Tiab] OR capybaras[Tiab])
#3 Reproductive/developmental toxicity terms	developmental biology [mh] OR developmental biology [tiab] OR embryonic and fetal development [mh] OR (embryonic [tiab] OR embryonically [tiab]) OR fetal development [tiab] OR growth and development [mh] OR growth and development [subheading] OR (development [tiab] OR developmental [tiab] OR developmentally [tiab]) OR embryology [mh] OR embryology [tiab] OR ecotoxicology [mh] OR ecotoxicology [tiab] OR ecology [mh] OR (ecology [tiab] OR ecological [tiab] OR ecologically [tiab]) OR toxicology [mh] OR (toxicology [tiab] OR toxicological [tiab] OR toxicologically [tiab]) OR toxicogenetics [mh] OR (toxicogenetic [tiab] OR toxicogenetics [tiab]) OR growth [mh] OR growth [tiab] OR environmental pollutants [mh] OR (environmental pollutant [tiab] OR environmental pollutants [tiab]) OR body weight [mh] OR (weight [tiab] OR weights [tiab] OR weighed [tiab]) OR embryo loss [mh] OR (embryo loss [tiab] OR embryo losses [tiab]) OR fetal resorption [mh] OR (fetal resorption [tiab] OR fetal resorptions [tiab]) OR gestational age [mh] OR (gestational age [tiab] OR gestational ages [tiab]) OR litter size [mh] OR (litter size [tiab] OR litter sizes [tiab]) OR endocrine disruptors [mh] OR (endocrine disruptor [tiab] OR endocrine disruptors [tiab] OR endocrine disruption [tiab]) OR reproduction [mh] OR reproduction [tiab] OR toxicity [subheading] OR toxicity [tiab] OR (toxic [tiab] OR toxics [tiab])
#4	**#1 AND #2 AND #3**

Table S2: Web of Science search terms.

Search	Web of Science
#1 Substance terms	TS=((perfluorooctanoic acid OR perfluorooctanoic acids) OR (perfluoroctanoic acid OR perfluoroctanoic acids) OR (perfluoro-n-octanoic acid OR perfluoro-n-octanoic acids) OR (pentadecafluorooctanoic acid OR pentadecafluorooctanoic acids) OR APFO OR (perfluorinated AND octanoic acid) OR (perfluorinated AND octanoic acids) OR (perfluorooctanoate OR perfluorooctanoates) OR perfluorooctanoyl chloride OR PFOA OR (fluorinated telomer alcohol OR fluorinated telomer alcohols) OR (fluoro-telomer alcohol OR fluoro-telomer alcohols) OR (fluorocarbon emulsion OR fluorocarbon emulsions) OR (perfluorocarbon OR perfluorocarbons) OR (fluorocarbon polymer OR fluorocarbon polymers) OR (fluorinated polymer OR fluorinated polymers) OR (octanoic acid OR octanoic acids) OR (caprylate OR caprylates) OR (polyfluoroalkyl OR polyfluoroalkyls OR polyfluoroalkylated) OR PFAA OR (perfluoroalkyl chemical OR perfluoroalkyl chemicals) OR (c8 AND perfluorinated) OR (fluoropolymer OR fluoropolymers OR fluoropolymeric) OR (fluorosurfactant OR fluorosurfactants) OR (perfluorochemical OR perfluorochemicals) OR PFCs OR (perfluoroalkyl carboxylate OR perfluoroalkyl carboxylates) OR (perfluorocarboxylate OR perfluorocarboxylates) OR PFCA OR (perfluorinated carboxylic acid OR perfluorinated carboxylic acids) OR FC 143 OR (pentadecafluorooctanoate OR pentadecafluorooctanoates))
#2 Experimental animal terms (modified from Hooijmans et al. 2010)	TS=(animals OR animal OR mice OR mus OR mouse OR murine OR woodmouse OR rats OR rat OR murinae OR muridae OR cottonrat OR cottonrats OR hamster OR hamsters OR cricetinae OR rodentia OR rodent OR rodents OR pigs OR pig OR swine OR swines OR piglets OR piglet OR boar OR boars OR "sus scrofa" OR ferrets OR ferret OR polecat OR polecats OR "mustela putorius" OR "guinea pigs" OR "guinea pig" OR cavia OR callithrix OR marmoset OR marmosets OR cebuella OR hapale OR octodon OR chinchilla OR chinchillas OR gerbillinae OR gerbil OR gerbils OR jird OR jirds OR merione OR meriones OR rabbits OR rabbit OR hares OR hare OR diptera OR flies OR fly OR dipteral OR drosphila OR drosophilidae OR cats OR cat OR carus OR felis OR nematoda OR nematode OR nematodes OR sipunculida OR dogs OR dog OR canine OR canines OR canis OR sheep OR sheeps OR mouflon OR mouflons OR ovis OR goats OR goat OR capra OR capras OR rupicapra OR chamois OR haplorhini OR monkey OR monkeys OR anthropoidea OR anthropoids OR saguinus OR tamarin OR tamarins OR leontopithecus OR hominidae OR ape OR apes OR pan OR paniscus OR "pan paniscus" OR bonobo OR bonobos OR troglodytes OR "pan troglodytes" OR gibbon OR gibbons OR siamang OR siamangs OR nomascus OR symphalangus OR chimpanzee OR chimpanzees OR prosimians OR "bush baby" OR prosimian OR bush babies OR galagos OR galago OR pongidae OR gorilla OR gorillas OR pongo OR pygmaeus OR "pongo pygmaeus" OR orangutans OR lemur OR lemurs OR lemuridae OR horse OR horses OR equus OR cow OR calf OR bull OR chicken OR chickens OR gallus OR quail OR bird OR birds OR quails OR poultry OR poultries OR fowl OR fowls OR reptile OR reptilia OR reptiles OR snakes OR snake OR lizard OR lizards OR alligator OR alligators OR crocodile OR crocodiles OR turtle OR turtles OR amphibian OR amphibians OR amphibia OR frog OR frogs OR bombina OR salientia OR toad OR toads OR "epidalea calamita" OR salamander OR salamanders OR eel OR eels OR fish OR fishes OR pisces OR catfish OR catfishes OR siluriformes OR arius OR heteropneustes OR sheatfish OR perch OR perches OR percidae OR perca OR trout OR trouts OR char OR chars OR salvelinus OR "fathead minnow" OR minnow OR cyprinidae OR carps OR carp OR zebrafish OR zebrafishes OR goldfish OR goldfishes OR guppy OR guppies OR chub OR chubs OR tinca OR barbels OR barbus OR pimephales OR promelas OR "poecilia reticulata" OR mullet OR mullets OR seahorse OR seahorses OR mugil curema OR atlantic cod OR shark OR sharks OR catshark OR anguilla OR salmonid OR salmonids OR whitefish OR whitefishes OR salmon OR salmons OR sole OR solea OR "sea lamprey" OR lamprey OR

Search	Web of Science
	lampreys OR pumpkinseed OR sunfish OR sunfishes OR tilapia OR tilapias OR turbot OR turbots OR flatfish OR flatfishes OR sciuridae OR squirrel OR squirrels OR chipmunk OR chipmunks OR suslik OR susliks OR vole OR voles OR lemming OR lemmings OR muskrat OR muskrats OR lemmus OR otter OR otters OR marten OR martens OR martes OR weasel OR badger OR badgers OR ermine OR mink OR minks OR sable OR sables OR gulo OR gulos OR wolverine OR wolverines OR mustela OR llama OR llamas OR alpaca OR alpacas OR camelid OR camelids OR guanaco OR guanacos OR chiroptera OR chiropteras OR bat OR bats OR fox OR foxes OR iguana OR iguanas OR xenopus laevis OR parakeet OR parakeets OR parrot OR parrots OR donkey OR donkeys OR mule OR mules OR zebra OR zebras OR shrew OR shrews OR bison OR bisons OR buffalo OR buffaloes OR deer OR deers OR bear OR bears OR panda OR pandas OR "wild hog" OR "wild boar" OR fitchew OR fitch OR beaver OR beavers OR jerboa OR jerboas OR capybara OR capybaras)
#3 Reproductive/developmental toxicity terms	TS=(developmental biology OR (embryonic OR embryonically) OR (development OR developmental OR developmentally) OR embryology OR ecotoxicology OR (ecology OR ecological OR ecologically) OR (toxicology OR toxicological OR toxicologically) OR (toxicogenetic OR toxicogenetics) OR growth OR (environmental pollutant OR environmental pollutants) OR (weight OR weights OR weighed) OR (embryo loss OR embryo losses) OR (fetal resorption OR fetal resorptions) OR (gestational age OR gestational ages) OR (litter size OR litter sizes) OR (endocrine disruptor OR endocrine disruptors OR endocrine disruption) OR reproduction OR (toxicity OR toxic OR toxics))
#4	**#1 AND #2 AND #3**

Table S3: Search strategy

Source[a]	Hits[b]
PubMed	1462
Web of Science (Thompson Reuters)	1060
Agency for Toxic Substances and Disease Registry (ATSDR) Interaction Profiles and Toxicological Profiles	1
Developmental and Reproductive Toxicology Database (DART)	10
EPA Science Inventory	88
USEPA Health and Environmental Studies Online (HERO)	52
National Institute for Occupational Safety and Health publications database (NIOSHTIC-2)	7
Toxicology Literature Online (TOXLINE)	85
Toxic Substances Control Act Test Submissions (TSCATS)	2

[a]Table presents sources for which search results were returned; sources that did not return search results follow: CalEPA Office of Environmental Health Hazard Assessment Risk Assessment; Chem IDplus Advanced; Chemspider; Chemical Carcinogenesis Research Information System (CCRIS); EPA Acute Exposure Guideline Levels Chemicals; EPA Integrated Risk Information System (IRIS); EPA National Environmental Publications Internet Site (NEPIS); EPA National Service Center for Environmental Publications (NSCEP); EPA Substance Registry Services; Environmental Mutagen Information Center (EMIC); European Chemicals Agency; Genetic Toxicology Data Bank (GENE-TOX); Health Canada First Priority Substances List (PSL1) Assessments; Health Canada Second Priority Substances List (PSL2) Assessments; Hazardous Substances Data Bank (HSDB); IARC Monographs on the Evaluation of Carcinogenic Risks to Humans; International Life Sciences Institute (ILSI); International Programme on Chemical Safety (IPCS); International Toxicity Estimates for Risk (ITER); US National Toxicology Program Management Status Report; US National Toxicology Program Results and Status Search; US National Toxicology Program Report on Carcinogens; Toxicology Excellence for Risk Assessment; Toxicology Data Network (TOXNET); NIOSH Registry of Toxic Effects of Chemical Substances (RTECS);WHO Concise International Chemical Assessment Documents; WHO Environmental Health Criteria. For additional information, see review protocol. [b]PubMed and Web of Science searches were performed on February 3, 2012; all remaining database searches were performed January 23, 2012-February 6, 2012

Table S4: Characteristics of Hu et al. 2010 (study ID 68).

Study Element	Description
Methods	Mouse developmental toxicological and immunotoxicological study
Participants	Wild-type C57Bl/6 mice Timed-pregnant GD6 animals obtained from supplier Total number of dams allocated: 48
Exposure	Dams treated with PFOA, ammonium salt (CAS# 3825-26-1) in drinking water (a*d libitum* access) from GD 6 to GD 17. Exposure groups: • 2 dose groups = 0.05, 1 mg PFOA/kg body weight/day; 16 dams each • 1 control group = non-treated water; 16 dams
Outcomes	1. Birth weight (g) - individual pups weighed at PND2 (birth occurred previous day or night). Number of dams analyzed: • 10 for each exposure group Not included in review: dam and offspring weight gain; organ weights; IgM and IgG antibody titers; serum PFOA concentrations
Notes	Author responded to requests for additional information. Raw data provided by study author. Litter sizes were statistically equal across doses and control groups.

Table S5: Characteristics of Yahia et al. 2010 (study ID 103).

Study Element	Description
Methods	Mouse developmental and reproductive toxicological study
Participants	Wild-type ICR (CD-1) mice In-house breeding protocol Total number of dams allocated: Unclear
Exposure	Experimental groups: • Prenatal time point: dams treated with PFOA via daily gavage from GD0 to GD17. • Postnatal time point: dams treated with PFOA via daily gavage from GD0 to GD18. Exposure groups: • 3 dose groups = 1, 5, 10 mg PFOA/kg body weight/day • 1 control group = deionized water
Outcomes	1. Fetal weight (g) – individual fetuses weighed at GD18. Number of dams analyzed: • 7, 9, 5, 8 for control, 1, 5, 10 mg/kg PFOA groups, respectively 2. Neonatal weight (g) – unclear at what time point pups were weighed. Number of dams analyzed: • 5 for each exposure group Not included in review: maternal organ effects; effects on maternal serum biochemical parameters; maternal weight; survival of offspring
Notes	Authors did not respond to requests for additional information. Failure to report pertinent study details resulted in exclusion from quantitative analysis for birth weight outcome, despite meeting inclusion criteria otherwise. Authors did not specify chemical form of PFOA (assume CAS# 335-67-1). For prenatal time point, no significant effects on survival. For postnatal time point, delayed delivery, high rate of stillborn birth (58%), and remaining live pups died within 6 hours after birth for 10 mg/kg PFOA group; 16% of neonates died in 5 mg/kg PFOA group.

Table S6: Characteristics of Hines et al. 2009 (study ID 260).

Study Element	Description
Methods	Mouse developmental toxicological study
Participants	Wild-type CD-1 mice Timed-pregnant GD0 animals obtained from supplier Total number of dams allocated: 85
Exposure	Dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily gavage from GD1 to GD17. Exposure groups: • 5 dose groups = 0.01, 0.1, 0.3, 1, 5 mg PFOA/kg body weight/day; 15, 15, 15, 15, 10 dams for 0.01, 0.1, 0.3, 1, 5 mg PFOA/kg body weight/day groups, respectively. • 1 control group = distilled water; 15 dams
Outcomes	1. Birth weight (g) – individual pups weighed at PND1 (day of birth or birth occurred previous night). Number of dams analyzed: • 10, 15, 15, 13, 14, 8 for control, 0.01, 0.1, 0.3, 1, 5 mg/kg PFOA groups, respectively Not included in review: maternal weight; offspring weight gain; offspring glucose, leptin, and insulin levels; offspring late life organ and body weight; offspring fat to lean ratio; offspring feed consumption; effect of ovariectomy
Notes	Author responded to requests for additional information. Raw data provided by study author. Litter sizes were statistically equal across doses and control groups.

Table S7: Characteristics of Fenton et al. 2009 (study ID 264).

Study Element	Description
Methods	Mouse pharmacokinetic study
Participants	Wild-type CD-1 mice Timed-pregnant GD14 animals obtained from supplier Total number of dams allocated: 100
Exposure	Experimental groups: • Prenatal time point: dams treated with PFOA, ammonium salt (CAS# 3825-26-1), via single gavage on GD17. • Postnatal time point: dams treated with PFOA, ammonium salt (CAS# 3825-26-1), via single gavage on GD17. Exposure groups: • 3 dose groups = 0.1, 1, 5 mg PFOA/kg body weight Prenatal time point: 5 dams/dose Postnatal time point: 5 dams/dose • 1 control group = deionized water Prenatal time point: 5 dams Postnatal time point: 5 dams
Outcomes	1. Fetal weight (g) – one individual fetus from each litter weighed at GD18. Number of dams analyzed: • 5, 5, 5, 4 for control, 0.1, 1, 5 mg/kg PFOA groups, respectively 2. Birth weight (g) – one individual pup from each litter weighed at PND1 (day of birth or birth occurred previous night). Number of dams analyzed: • 5, 4, 4, 5 for control, 0.1, 1, 5 mg/kg PFOA groups, respectively Not included in review: PFOA serum concentration prior to birth; dam PFOA serum concentrations; pup PFOA serum concentrations
Notes	Author responded to requests for additional information. Raw data provided by study author. Number of live fetuses and litter sizes at birth were statistically equal across doses and control groups.

Table S8: Characteristics of White et al. 2009 (study ID 312).

Study Element	Description
Methods	Mouse developmental toxicological study
Participants	Wild-type CD-1 mice Timed-pregnant GD0 animals obtained from supplier Total number of dams allocated: 112
Exposure	Dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily gavage from GD8 to GD17. At birth, pups were cross-fostered to obtain the following groups: 1) never exposed, 2) exposed in utero and via lactation, 3) exposed only in utero, 4) exposed only via lactation. Exposure groups: • 1 dose groups = 5 mg PFOA/kg body weight/day; 56 dams • 1 control group = deionized water; 56 dams
Outcomes	1. Birth weight (g) – 3 individual female pups from each litter weighed at PND1 (at least 12 hours after birth and cross-foster); only 2 groups of pups relevant for review: 1) never exposed, 2) exposed in utero and via lactation. Number of dams analyzed: • 4 for each exposure group Not included in review: 2 additional studies - late-life effects cross-foster study and restricted-exposure study; mammary gland development scores; circulating serum PFOA concentration; mammary gland differentiation in offspring; serum PFOA dosimetry
Notes	Author responded to requests for additional information. Raw data provided by study author. Litter sizes were statistically equal across doses and control groups.

Table S9: Characteristics of Abbott et al. 2007 (study ID 528).

Study Element	Description
Methods	Mouse developmental toxicological study
Participants	Wild-type 129S1/SvlmJ mice In-house breeding protocol Total number of dams allocated: 157
Exposure	Dams treated with PFOA, ammonium salt (CAS# 3825-26-1), via daily gavage from GD1 to GD17. Exposure groups: • 7 dose groups = 0.1, 0.3, 0.6, 1, 5, 10, 20 mg PFOA/kg body weight/day; 15, 16, 26, 25, 12, 11, 16 dams for 0.1, 0.3, 0.6, 1, 5, 10, 20 mg PFOA/kg body weight/day groups, respectively • 1 control group = deionized water; 36 dams
Outcomes	1. Birth weight (g) – pups grouped by sex and weighed at PND1 (day of birth or birth occurred previous night). Number of dams analyzed: • 19, 10, 7, 9, 13 for 0.1, 0.3, 0.6, 1 mg/kg PFOA groups, respectively Not included in review: studies of PPARα knockout mouse model; maternal weight and reproductive outcomes; liver weight in dams and offspring; offspring survival, development, and growth; serum PFOA in dams and offspring
Notes	Author responded to requests for additional information. Raw data provided by study author. Study authors noted that the 129S1/SvlmJ strain is more sensitive to PFOA exposure than other strains, such as CD-1 strain mice. Litter sizes were statistically equal across dose groups up to 1 mg/kg PFOA and control groups; incidence of full litter resorption was statistically significantly higher in 5 mg/kg PFOA groups (83, 80, 100% of dams had full litter resorption for 5, 10, 20 mg/kg PFOA groups, respectively).

Table S10: Characteristics of White et al. 2007 (study ID 566).

Study Element	Description
Methods	Mouse developmental toxicological study
Participants	Wild-type CD-1 mice Timed-pregnant GD0 animals obtained from supplier Total number of dams allocated: 60
Exposure	Dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily gavage for following time periods during pregnancy: GD1-17, GD8-17, GD12-17. Exposure groups: • 1 dose groups = 5 mg PFOA/kg body weight/day; 14, 16, 16 dams for GD1-17, GD8-17, GD12-17 time periods, respectively • 1 control group = deionized water; 14 dams for GD1-17 time period
Outcomes	1. Birth weight (g) – individual pups weighed at PND1 (day of birth or birth occurred previous night). Number of dams analyzed: • 5 mg /kg PFOA: 6, 11, 10 for GD1-17, GD8-17, GD12-17 time periods, respectively • Control: 10 for GD1-17 time periods Not included in review: maternal weight gain; dam lactating mammary gland development; milk protein gene expression; blood PFOA concentrations; offspring mammary gland development.
Notes	Author responded to requests for additional information. Raw data provided by study author. Litter sizes and number of uterine implantation sites were statistically equal across doses and control groups.

Table S11: Characteristics of Wolf et al. 2007 (study ID 571).

Study Element	Description
Methods	Mouse developmental toxicological study
Participants	Wild-type CD-1 mice Timed-pregnant GD0 animals obtained from supplier Total number of dams allocated: 182
Exposure	Experimental Groups • Cross-foster: dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily gavage from GD1-GD17. • Windows of sensitivity: dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily gavage for following time periods during pregnancy: GD7-17, GD10-17, GD13-17, GD15-17. Exposure groups: • 3 dose groups = 3, 5, 20 mg PFOA/kg body weight/day Cross-foster: 28, 36 dams for 3, 5 mg PFOA/kg body weight/day groups, respectively Windows of sensitivity: 5 mg PFOA/kg body weight/day: 14, 14, 12, 12 dams for GD7-17, GD10-17, GD13-17, GD15-17 time periods, respectively 20 mg PFOA/kg body weight/day: 6 dams for GD15-17 time period • 1 control group = deionized water Cross-foster: 48 dams Windows of sensitivity: 12 dams for GD7-17 time period
Outcomes	1. Birth weight (g) – pups grouped by sex and weighed at birth (birth monitored at time intervals throughout night). Number of dams analyzed: • Control: 38, 7 for GD1-17, GD7-17 time points, respectively • 3 mg/kg PFOA: 24 for GD1-17 time point • 5 mg/kg PFOA: 25, 13, 13, 10, 10 for GD1-17, GD7-17, GD10-17, GD13-17, GD15-17 time points, respectively • 20 mg/kg PFOA: 3 for GD15-17 time point Not included in review: maternal weight; dam reproductive outcomes; dam liver weight; foster offspring survival, development, and growth; dam and offspring serum PFOA levels.
Notes	Author responded to requests for additional information. Raw data provided by study author. Litter sizes were statistically equal across doses and control groups. Incidence of whole litter loss was statistically significantly increased in the 5 mg/kg PFOA group as compared to other treatment groups (not caused by litter resorptions as statistically equal number of uterine implantation sites as compared to control).

Table S12: Characteristics of Lau et al. 2006 (study ID 635).

Study Element	Description
Methods	Mouse developmental toxicological study
Participants	Wild-type CD-1 mice Timed-pregnant GD0 animals obtained from supplier Total number of dams allocated: Unclear
Exposure	Dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily gavage from GD1-GD17. Exposure groups: • 6 dose groups = 1, 3, 5, 10, 20, 40 mg PFOA/kg body weight/day • 1 control group = water
Outcomes	1. Fetal weight (g) – fetuses weighed individually at GD18. Number of dams analyzed: • 45, 17, 17, 27, 26, 42, 9 dams for 0, 1, 3, 5, 10, 20, 40 mg/kg PFOA groups, respectively. 2. Birth weight (g) – pups weighed as litter at birth (birth monitored hourly). Number of dams analyzed: • 24, 8, 8, 30, 26, 7 for 0, 1, 3, 5, 10, 20 mg/kg PFOA groups, respectively. Not included in review: serum PFOA compared for rats and mice; maternal weight gain; dam reproductive outcomes; fetal teratology; effect on time to parturition; offspring survival; offspring weight gain; offspring developmental landmarks; benchmark dose estimates.
Notes	Author responded to requests for additional information. Raw data provided by study author. Full litter resorptions were statistically significantly increased in 5mg/kg PFOA and higher doses (100% litter resorption in 40 mg/kg PFOA group). Litter size at birth was statistically significantly decreased in 20 mg/kg PFOA group.

Table S13: Characteristics of Hinderliter et al. 2005 (study ID 711).

Study Element	Description
Methods	Rat pharmacokinetic study
Participants	Wild-type Sprague-Dawley rats Timed-pregnant GD1 animals obtained from supplier Total number of dams allocated: 20
Exposure	Dams treated with PFOA, ammonium salt (CAS# 3825-26-1), via daily gavage from GD4-GD21. Exposure groups: • 3 dose groups = 3, 10, 30 mg PFOA/kg body weight/day; 20 dams for each dose group • 1 control group = deionized water; 20 dams
Outcomes	1. Birth weight (g) – pups grouped by sex and weighed at PND1 (day when birth complete). Number of dams analyzed: • 5 dams for each dose group Not included in review: mortality; dam weight gain; offspring growth and survival; dam and offspring PFOA concentration
Notes	Author responded to requests for additional information. Published study does not present data on this outcome; authors provided full industry report performed according to good laboratory practices (GLP) that included data for birth weight. Litter sizes were statistically equal across doses and control groups.

Table S14: Characteristics of Staples et al. 1984 (study ID 1871).

Study Element	Description
Methods	Rat developmental toxicological study
Participants	Wild-type Sprague-Dawley rats In-house breeding protocol Total number of dams allocated: 224
Exposure	Experimental groups: • Inhalation: dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily inhalation as a dust for 6 hour period from GD6-15; prenatal and postnatal time points. • Gavage: dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily gavage from GD6-15; prenatal and postnatal time points. Exposure groups: • 4 inhalation dose groups = 0.1, 1, 10, 25 mg PFOA/m^3 Prenatal time point: 24, 24, 15, 12 dams for 0.1, 1, 10, 25 mg PFOA/m^3 groups, respectively Postnatal time point: 12, 12, 6, 12 dams for 0.1, 1, 10, 25 mg PFOA/m^3 groups, respectively • 3 inhalation control groups = in-house air only, in-house air pair-fed to 10 mg/m^3 dose group, in-house air pair-fed to 25 mg/m^3 dose group Prenatal time point: 24, 6, 6 dams for in-house air only, pair-fed to 10 mg/m^3, and pair-fed to 25 mg/m3 groups, respectively Postnatal time point: 18 dams for in-house air only • 1 gavage dose group = 100 mg PFOA/kg body weight/day Prenatal time point: 25 dams Postnatal time point: 12 dams • 1 gavage control group = stripped corn oil Prenatal time point: 25 dams Postnatal time point: 12 dams
Outcomes	1. Fetal weight (g) – individual fetuses weighed at GD21. Number of dams analyzed: • Inhalation: 23, 24, 23, 15, 7, 6, 5 for 0, 0.1, 1, 10, 25 mg PFOA/m^3, pair-fed to 10 mg/m^3, pair-fed to 25 mg/m^3 groups, respectively. • Gavage: 24, 22 for 0, 100 mg/kg PFOA groups, respectively. 2. Birth weight (g) – individual pups weighed at PND1 (day of birth). Number of dams analyzed: • Inhalation: 18, 10, 11, 6, 9 for 0, 0.1, 1, 10, 25 mg PFOA/m^3 groups, respectively. • Gavage: 12, 9 for 0, 100 mg/kg PFOA groups, respectively. Not included in review: fetal teratology; dam body and liver weights; offspring survival; dam survival; offspring defects.
Notes	Study authors contacted to provide additional information, but information could not be obtained due to length of time since study conducted. Incidence of litter resorptions and litter size at birth were statistically equal across doses and control groups.

Table S15: Characteristics of Boberg et al. 2008 (study ID 3061).

Study Element	Description
Methods	Rat metabolic developmental toxicological study
Participants	Wild-type Wistar rats Timed-pregnant GD3 animals obtained from supplier Total number of dams allocated: 18
Exposure	Dams treated with PFOA (CAS# 333-67-1) via daily gavage from GD7-GD20/21. Exposure groups: • 1 dose groups = 20 mg PFOA/kg body weight/day; 8 dams • 1 control group = corn oil; 10 dams
Outcomes	1. Fetal weight (g) – fetuses weighed individually at GD20/21. Number of dams analyzed: • 5, 6 for 0, 20 mg/kg PFOA groups, respectively. Not included in review: outcomes for treatment with diisobutyl phthalate and butylparaben; steroid hormone measurement; plasma levels of metabolic chemicals; mRNA expression; P450c17 and PPARγ protein levels in testes.
Notes	Author responded to requests for additional information. Raw data provided by study author; published study does not present data on this outcome. Author noted that "some animals were sacrificed one day too early for their age. Therefore, in the GD21 group, about one fourth are GD20 and three fourths are GD21". Study did not discuss fetal mortality.

Table S16: Characteristics of Onishchenko et al. 2011 (study ID 3610).

Study Element	Description
Methods	Mouse neurobehavioral developmental toxicological study
Participants	Wild-type C57BL/6 mice In-house breeding protocol Total number of dams allocated: 16
Exposure	Dams treated with PFOA applied to food from GD1-GD20. Exposure groups: • 1 dose groups = 0.3 mg PFOA/kg body weight/day; 10 dams • 1 control group = food applied with ethanol; 6 dams
Outcomes	1. Birth weight (g) – pups weighed at PND 1 (not clearly defined). Number of dams analyzed: • 6, 9 for 0, 0.3 mg/kg PFOA groups, respectively. Not included in review: outcomes from treatment with PFOS; PFOA concentrations in tissues; locomotor and exploratory activity; circadian activity; anxiety-related behavior; depression-like behavior; muscle strength; motor coordination
Notes	Author responded to requests for additional information. Raw data provided by study author; published study does not present data on this outcome. The author reported results for a larger number of PFOA-treated dams than originally allocated in the paper. Results from the study author do not clearly explain if the body weights are litter averages. Failure to report pertinent study details resulted in exclusion from quantitative analysis for birth weight outcome, despite meeting inclusion criteria otherwise. Authors did not specify chemical form of PFOA (assume CAS# 335-67-1). Litter sizes were statistically equal across doses and control groups.

Table S17: Characteristics of White et al. 2011 (study ID 3862).

Study Element	Description
Methods	Mouse multigenerational developmental toxicological study
Participants	Wild-type CD-1 mice Timed-pregnant GD0 animals obtained from supplier Total number of dams allocated: 33
Exposure	Dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily gavage from GD1 to GD17. Exposure groups: • 2 dose groups = 1, 5 mg PFOA/kg body weight/day; 12, 11 dams for 1, 5 mg PFOA/kg body weight/day groups, respectively. • 1 control group = deionized water; 10 dams
Outcomes	1. Birth weight (g) – F1 pups grouped by sex and weighed at PND1 (day of birth or birth occurred previous night). Number of dams analyzed: • 18, 23, 19 dams for 0, 1, 5 mg/kg PFOA groups, respectively. Not included in review: results for F2 generation; dam weight gain; offspring growth and survival; mammary development; lactation success; water consumption; serum PFOA
Notes	Author responded to requests for additional information. Raw data provided by study author; published study does not present data on this outcome. Published study included groups treated with drinking water containing PFOA, but study authors did not provide data for these studies and stated that only the groups treated via gavage could be compared. The number of uterine implantation sites was statistically equal across doses and control groups. Litter size at birth and prenatal survival were statistically significantly decreased in 5 mg/kg PFOA groups.

Table S18: Characteristics of York 2002 (study ID 5122).

Study Element	Description
Methods	Rat multigenerational developmental toxicological study
Participants	Wild-type Sprague-Dawley rats Virgin animals obtained from supplier Total number of dams allocated: 150
Exposure	Dams treated with PFOA, ammonium salt (CAS# 3825-26-1) via daily gavage from 70 days prior to breeding through lactation. Exposure groups: • 4 dose groups = 1, 3, 10, 30 mg PFOA/kg body weight/day; 30 dams for each dose group • 1 control group = deionized water; 30 dams
Outcomes	1. Birth weight (g) – F1 pups weighed individually at PND1 (day of birth). Number of dams analyzed: • 28, 27, 29, 29, 28 for 0, 1, 3, 10, 30 mg/kg PFOA groups, respectively Not included in review: duration of gestation; fertility index; number/sex offspring per litter; number of implantation sites; maternal behavior; necropsy for gross lesions; organ weight and evaluation
Notes	Author responded to requests for additional information. This study was an industry report performed according to good laboratory practices (GLP). The report supports papers in peer-reviewed literature. Litter sizes at birth were statistically equal across doses and control groups.

Table S19: Characteristics of Hagenaars et al. 2011 (study ID 59).

Study Element	Description
Methods	Zebrafish (*Danio rerio*) developmental toxicological study
Participants	Wild-type zebrafish eggs In-house breeding protocol Eggs collected 30 min after spawning Total number of eggs allocated: 368
Exposure	Eggs immersed in PFOA (CAS# 335-67-1) test solutions within 60 min after spawning and continuously treated until 120 hours post fertilization (hpf). Hatching typically occurs at 72 hpf. Exposure Groups: • 8 dose groups = 15, 20, 30, 40, 50, 75, 100, 250 mg PFOA/L freshwater; 40 eggs each • 1 control group = freshwater; 48 eggs
Outcomes	1. Length (mm) - individual zebrafish embryos measured at 120 hpf. Number of zebrafish embryos analyzed: • 34, 36, 32, 33, 30, 33, 33, 31, 30 for control, 15, 20, 30, 40, 50, 75, 100, 250 mg PFOA/L freshwater groups, respectively Not included in review: outcomes for treatment with PFOS, PFBS, and PFBA; concentration-response relationship; malformations; hatching rate; heart rate; correlation heart rate and length.
Notes	Author responded to requests for additional information. Raw data provided by study author (including 15, 20 mg PFOA/L doses not presented in paper). PFOA induced mortality at highest doses in dose-dependent manner (calculated EC_{50} 205.72 mg PFOA/L). Embryo hatching statistically significantly delayed in 50 mg PFOA/L and higher dose groups.

Table S20: Characteristics of Wang et al. 2010 (study ID 86).

Study Element	Description
Methods	Fruit fly (*Drosophila melanogaster)* developmental toxicological study
Participants	Wild-type flies (W118 stock) In-house breeding protocol Total number of female flies allocated (larvae stage): 300 Total number of female flies allocated (pupae stage): Unclear
Exposure	Experimental groups: • Larvae: one-day-old female flies allowed to lay eggs for 2 hours in vials containing food with PFOA, ammonium salt (CAS# 3825-26-1). Eggs hatched and developed on media and larvae were collected at 30, 48, 72, 96, and 110 hours after egg laying (AEL). • Pupae: same regimen as above. Larvae allowed to develop to white pupae stage. Exposure groups: • 2 dose groups = 100, 500 μM PFOA in food Larvae group: 20 female flies/dose, for each time point Pupae group: Unclear how many female flies originally allocated • 1 control group = untreated food Larvae group: 20 female flies for each time point Pupae group: Unclear how many female flies originally allocated
Outcomes	1. Larval volume (arbitrary units) – calculated based on individual measurements of length and diameter. Number of larvae analyzed: • Control: 10, 23, 30, 31, 28 for 30, 48, 72, 96, 110 AEL groups, respectively • 100 μM PFOA in food: 15, 20, 26, 33, 33 for 30, 48, 72, 96, 110 AEL groups, respectively • 500 μM PFOA in food: 14, 15, 46, 25, 29 for 30, 48, 72, 96, 110 AEL groups , respectively 2. Pupae weight (g) – individual pupae weights. Number of pupae analyzed: • 38, 35, 25 for control, 100, 500 μM PFOA in food groups, respectively Not included in review: lifespan, behavior, numbers of emerging progeny, lethality, developmental progress, effects of nutrient supplementation
Notes	Author responded to requests for additional information. Raw data provided by study author (including length measurements plotted for analysis). Emergence of progeny was statistically significantly reduced in both PFOA dose groups (24% and 73% decrease compared to control group for 100, 500 μM PFOA in food groups, respectively).

Table S21: Characteristics of Pinkas et al. 2010 (study ID 187).

Study Element	Description
Methods	Chicken neurodevelopmental toxicological study
Participants	Cobb I chicken broiler strain (*Gallus gallus domesticus*) Fertile heterogeneous stock eggs obtained from supplier Total number of eggs allocated: Unclear
Exposure	Eggs injected with PFOA (CAS# 335-67-1) at incubation day 0. Exposure groups: • 2 dose groups = 5, 10 mg PFOA/kg egg • 1 control group = saline
Outcomes	1. Hatchling weight (g) – individual hatchlings weighed 24 hours after hatching. Number of hatchlings analyzed: • 30, 12, 10 for control, 5, 10 mg PFOA/kg egg groups, respectively Not included in review: outcomes from treatment with PFOS; hatching and survival; morphological and functional scores; imprinting scores; protein kinase C concentrations in the intermedial part of the hyperstriatum ventrale
Notes	Author responded to requests for additional information. Study author provided data estimates used to create figure in paper. On incubation day 19, survival of embryos was statistically significantly reduced (approx 45-55%) in the PFOA treated groups as compared to controls; hatching was statistically significantly reduced (70-80%) compared to controls.

Table S22: Characteristics of O'Brien et al. 2009 (study ID 236).

Study Element	Description
Methods	Chicken developmental toxicological study
Participants	White leghorn chicken (*Gallus gallus domesticus*) Eggs obtained from supplier Total number of eggs allocated: 120
Exposure	Eggs injected with PFOA at incubation day 0. Exposure groups: • 4 dose groups = 0.01, 0.1, 1, 10 mg PFOA/kg egg; 20 eggs for each dose group • 2 control groups = uninjected, DMSO; 20 eggs for each control group
Outcomes	1. Embryo weight (g) – individual embryos weighed at pipping star or day 22, whichever came first. Number of embryos analyzed: • 12, 18, 17, 15, 16, 15 for uninjected, DMSO only, 0.01, 0.1, 1, 10 mg PFOA/kg egg egg groups, respectively Not included in review: outcomes from treatment with PFUdA and PFDS; pipping success; hepatic PFC concentrations; mRNA expression
Notes	Author responded to requests for additional information. Raw data provided by study author; published study does not present data on this outcome. Authors did not specify chemical form of PFOA (assume CAS# 335-67-1). Pipping success and developmental stage at embryo death were statistically equal across doses and control groups.

Table S23: Characteristics of Jiang et al. 2012 (study ID 3926).

Study Element	Description
Methods	Chicken developmental toxicological study
Participants	Chicken (*Gallus gallus*) Fertile stock eggs obtained from supplier Total number of eggs allocated: 176
Exposure	Experimental groups: • Embryonic time point: eggs injected with PFOA at incubation day 0. • Hatchling time point: eggs injected with PFOA at incubation day 0. Exposure groups: • 3 dose groups = 0.5, 1, 2 mg PFOA/kg egg Embryonic time point: 10, 12, 9 eggs for 0.5, 1, 2 mg PFOA/kg egg groups, respectively. Hatchling time point: 22, 23, 24 eggs for 0.5, 1, 2 mg PFOA/kg egg groups, respectively. • 2 control groups = uninjected, sunflower oil Embryonic time point: 10, 9 eggs for uninjected, sunflower oil groups, respectively. Hatchling time point: 36, 21 eggs for uninjected, sunflower oil groups, respectively.
Outcomes	1. Yolk free body weight (g) Embyronic time point: individual embryos removed from eggs and weighed at embryonic day 19. Number of embryos analyzed: • 9, 8, 7, 9, 7 for uninjected, sunflower oil only, 0.5, 1, 2 mg PFOA/kg egg groups, respectively Hatchling time point: individual hatchlings weighed 16-24 hours after hatching. Number of hatchlings analyzed (same hatchlings used to examine crown to rump length outcome described below): • 26, 10, 11, 12, 9 for uninjected, sunflower oil only, 0.5, 1, 2 mg PFOA/kg egg groups, respectively 2. Crown to rump length (mm) Hatchling time point: individual hatchlings measured 16-24 hours after hatching - see above for number of hatchlings analyzed. Not included in review: embryo and hatchling heart weight; embryo and hatchling liver weight; embryo and hatchling mortality; hatchability; embryo cardiac morphology; hatchling cardiac ultrasound; hatchling cardiac myofibril ATPase; hatchling serum PFOA concentration
Notes	Author responded to requests for additional information. Raw data provided by study author. Authors did not specify chemical form of PFOA (assume CAS# 335-67-1). In embryos, mortality was statistically significantly increased in 2 mg PFOA/kg egg group (76% increase compared to control). In hatchlings, mortality and hatching were statistically equal across doses and control groups.

Table S24: Characteristics of Spachmo and Arukwe 2012 (study ID 3932).

Study Element	Description
Methods	Salmon endocrine and developmental toxicological study
Participants	Atlantic salmon (*Salmo salar*) Eggs obtained from supplier Total number of eggs allocated: Unclear
Exposure	Eggs exposed to PFOA in water through day 48. Hatching occurred at day 20. Larvae were collected at days 21, 35, 49, and 56. Exposure groups: • 1 dose groups = 100 μg PFOA/L water • 1 control group = water with carrier solvent (methanol)
Outcomes	1. Length (cm) – larvae measured using microscope with ruler and digital camera. Number of larvae analyzed (same larvae used to examine dry weight outcome described below): • 10 randomly selected larvae for each time point and dose group 2. Dry weight (g) – larvae dried in heat cabinet and weighed using a micro-weight scale - see above for number of larvae analyzed. Not included in review: outcomes from treatment with PFOS; bone development; effects on HTP-axis; effects on ER expression; effects on GH-IGF axis; effects on chondrogenic and osteogenic pathways
Notes	Authors did not respond to requests for additional information. Authors did not specify chemical form of PFOA (assume CAS# 335-67-1). Data estimates for figures presented in the published paper were obtained using an online digital ruler. Study did not discuss survival.

Table S25: Risk of bias summary of Hu et al. 2010 (study ID 68).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	Randomization scheme not discussed; adequate randomization is not standard protocol for these types of experiments.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Based on author response, dams and offspring were adequately followed; author reported numbers allocated.
Selective reporting	Low risk	Outcomes outlined in abstract/methods section of paper were reported in results section; "n" provided by author; raw data provided by author.
Conflict of interest	Probably low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. However, no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S26: Risk of bias summary of Yahia et al. 2010 (study ID 103).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	Randomization scheme not discussed; adequate randomization is not standard protocol for these types of experiments.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Probably high risk	Dams were not adequately followed; offspring mortality reported; numbers allocated not reported.
Selective reporting	Probably high risk	Neonatal weight outcome reported in results section of paper, but not pre-specified in methods section; time point of weight measurement not reported; adequate "n" reported for neonate outcomes; average values reported for fetal outcome so can calculate "n" analyzed.
Conflict of interest	Probably low risk	Funding source not stated, but associated persons appear to be from government and/or academia only and free of financial interests in study results. However, no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S27: Risk of bias summary of Hines et al. 2009 (study ID 260).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	Mice were "randomly distributed among treatment groups", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Probably low risk	Based on author response, dams were adequately followed; offspring mortality not reported, but noted that losses equivalent between groups; author reported numbers allocated.
Selective reporting	Probably low risk	Author reported different numbers allocated than reported in paper and noted that a figure in the paper displayed male and female weights even though the legend specifies female weights only; "n" provided by author; raw data provided by author.
Conflict of interest	Probably high risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. However, acknowledgement was made to employee from company financially invested in PFOA (Dow) for "constructive input on this manuscript" and no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S28: Risk of bias summary of Fenton et al. 2009 (study ID 264).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	Mice were "were weighed and randomly distributed among PFOA treatment groups", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Probably high risk	Based on author response, offspring mortality not assessed; dams were adequately followed; weighed one pup per litter; numbers allocated reported.
Selective reporting	Probably low risk	Author reported that one pup per litter included in weight outcome, but this was not clearly stated in paper; "n" provided by author; raw data provided by author.
Conflict of interest	Low risk	Funding source not stated, but associated persons appear to be from government and/or academia only and free of financial interests in study results. "The author declares that there are no conflicts of interest."
Other bias	Low risk	No other potential biases are suspected.

Table S29: Risk of bias summary of White et al. 2009 (study ID 312).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Upon arrival at the animal facility on GD 0, mice were weighed and randomly assigned to one of two treatment groups", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Probably high risk	Based on author response, offspring were adequately followed; only reported on results from 3 pups weighed from each of 4 dams (56 allocated/group); numbers allocated reported.
Selective reporting	Low risk	Outcomes outlined in abstract/methods section of paper were reported in results section; adequate "n" reported; raw data provided by author.
Conflict of interest	Probably high risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. "The authors declare that there are no conflicts of interest." However, company financially invested in PFOA (3M) provided analysis of PFOA chemical used in study.
Other bias	Low risk	No other potential biases are suspected.

Table S30: Risk of bias summary of Abbott et al. 2007 (study ID 528).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Plug positive female mice were weighed, randomly assigned to treatment groups", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Based on author response, dams and offspring were adequately followed; author reported numbers allocated.
Selective reporting	Low risk	Outcomes outlined in abstract/methods section of paper were reported in results section; "n" provided by author; raw data provided by author.
Conflict of interest	Probably low risk	Funding source not stated, but associated persons appear to be from government and/or academia only and free of financial interests in study results. However, no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S31: Risk of bias summary of White et al. 2007 (study ID 566).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Animals were weighed upon arrival and randomly distributed among four treatment groups", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Based on author response, dams and offspring were adequately followed; numbers allocated reported.
Selective reporting	Low risk	Outcomes outlined in abstract/methods section of paper were reported in results section; "n" provided by author; raw data provided by author.
Conflict of interest	Probably low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. However, no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S32: Risk of bias summary of Wolf et al. 2007 (study ID 571).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Upon arrival at the animal facility, mice were weighed and randomly assigned", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Dams and offspring were adequately followed; numbers allocated reported.
Selective reporting	Low risk	Outcomes outlined in abstract/methods section of paper were reported in results section; "n" provided by author; raw data provided by author.
Conflict of interest	Probably high risk	Funding source not stated, but associated persons appear to be from government and/or academia only and free of financial interests in study results. However, company financially invested in PFOA (3M) provided analysis of PFOA chemical used in study and no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S33: Risk of bias summary of Lau et al. 2006 (study ID 635).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	High risk	"Animals were randomly assigned to treatment groups". Author noted that mice were ranked according to weight at arrival then assigned evenly to each group at "random" but that a component such as a random number generator was not used in the sequence generation process.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Probably low risk	Dams and offspring were adequately followed; numbers allocated not reported.
Selective reporting	Probably low risk	Outcomes outlined in abstract/methods section of paper were reported in results section; adequate "n" provided by author for fetal and birth outcomes (pups per litter provided as range).
Conflict of interest	Probably high risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. However, company financially invested in PFOA (3M) provided analysis of PFOA chemical used in study and no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S34: Risk of bias summary of Hinderliter et al. 2005 (study ID 711).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Dams were ranked on… body weights and assigned to control and experimental groups by random sampling from the ranked list...Rats in each group were then randomly assigned to each subset", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Based on author response, dams and pups were adequately followed; numbers allocated reported.
Selective reporting	Low risk	Weight outcome outlined in methods section of paper, but not reported in results section; however, detailed outcome data presented in report provided by author; "n" provided by author; raw data provided by author.
Conflict of interest	High risk	Du Pont and 3M sponsored study and authors were employed by companies.
Other bias	Low risk	No other potential biases are suspected.

Table S35: Risk of bias summary of Staples et al. 1984 (study ID 1871).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	High risk	Females were ranked within breeding days by body weight and assigned to groups by rotating in order of rank.
Allocation concealment	High risk	Females were allocated by rotation.
Blinding	Probably low risk	"To limit possible bias in the examination of maternal and fetal specimens, the dams were coded (group designation unknown to examiner) from just before sacrifice until all maternal and fetal data were collected"; unclear if applies to birth outcomes.
Incomplete outcome data	Low risk	Dams and offspring were adequately followed; numbers allocated reported.
Selective reporting	Probably low risk	Outcomes outlined in abstract/methods section of paper were reported in results section; average values reported so can calculate "n" analyzed.
Conflict of interest	High risk	Du Pont sponsored study and authors were employed by the company.
Other bias	Low risk	No other potential biases are suspected.

Table S36: Risk of bias summary of Boberg et al. 2008 (study ID 3061).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"The dams were randomized into seven groups of eight with similar body weight distributions", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Based on author response, dams and pups were adequately followed; numbers allocated reported.
Selective reporting	Probably low risk	Author reported different numbers allocated than reported in paper; weight outcome pre-specified in methods section of paper and reported as "data not shown" in results section; however, detailed outcome data provided by author; "n" provided by author; raw data provided by author.
Conflict of interest	Low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. Authors claim "none" for conflicts of interest.
Other bias	Low risk	No other potential biases are suspected.

Table S37: Risk of bias summary of Onishchenko et al. 2011 (study ID 3610).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	Randomization scheme not discussed; adequate randomization is not standard protocol for these types of experiments.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Probably high risk	Dams and offspring were not adequately followed, but noted "litter size and sex ration were similar in control and exposed groups"; number allocated reported, but author provided data for more dams than allocated.
Selective reporting	Probably high risk	Weight outcome reported in results section of paper, but not pre-specified in methods section; "n" provided by author, but unclear if pup number or dam number and greater than numbers allocated in paper; author provided raw data.
Conflict of interest	Probably low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. However, no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S38: Risk of bias summary of White et al. 2011 (study ID 3862).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Timed pregnant dams were randomly distributed among five treatment groups", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Dams and offspring were adequately followed; numbers allocated reported, but author provided different numbers allocated.
Selective reporting	Probably low risk	Author reported different numbers allocated than reported in paper; weight outcome outlined in methods section paper, but not reported in results section; however, detailed outcome data provided by author; "n" provided by author; raw data provided by author.
Conflict of interest	Low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. "The authors declare they have no actual or potential competing financial interests."
Other bias	Low risk	No other potential biases are suspected.

Table S39: Risk of bias summary of York 2002 (study ID 5122).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Low risk	"Upon arrival parental generation rats will be assigned to individual housing on the basis on computer-generated random units… The rats will be assigned to dosage groups based on computer-generated (weight-ordered) randomization procedures."
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Dams and offspring were adequately followed; numbers allocated reported.
Selective reporting	Low risk	Outcomes outlined in abstract/methods section of paper were reported in results section; adequate "n" reported.
Conflict of interest	High risk	3M sponsored study and authors were employed by company.
Other bias	Low risk	No other potential biases are suspected.

Table S40: Risk of bias summary of Hagenaars et al. 2011 (study ID 59).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	Randomization scheme not discussed; adequate randomization is not standard protocol for these types of experiments.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Hatchability and mortality reported; numbers allocated reported.
Selective reporting	Probably low risk	Author reported different numbers allocated than reported in paper; "n" provided by author; raw data provided by author.
Conflict of interest	Probably low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. However, no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S41: Risk of bias summary of Wang et al. 2010 (study ID 86).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Cohorts of…flies were assigned randomly into vials"; unclear if applies to growth outcomes; details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Probably high risk	Hatchability and mortality not reported; only subset of treatment group analyzed for outcome; numbers allocated reported.
Selective reporting	Probably low risk	Pupae weight outcome reported in results section of paper, but not pre-specified in methods section; "n" provided by author; raw data provided by author.
Conflict of interest	Probably low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. However, no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S42: Risk of bias summary of Pinkas et al. 2010 (study ID 187).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	Randomization scheme not discussed; adequate randomization is not standard protocol for these types of experiments.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Probably low risk	Hatchlings were adequately followed; numbers allocated not reported.
Selective reporting	Probably high risk	Weight outcome reported in results section paper, but not pre-specified in methods section; adequate "n" reported.
Conflict of interest	Low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. "No author has any conflict of interest to disclose."
Other bias	Low risk	No other potential biases are suspected.

Table S43: Risk of bias summary of O'Brien et al. 2009 (study ID 236).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably low risk	"Eggs were randomly distributed among 4 dose groups", although details of random sequence generation were not discussed.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably high risk	Blinding not discussed; adequate blinding is not standard protocol for these types of experiments.
Incomplete outcome data	Low risk	Based on author response, embryos were adequately followed; numbers allocated reported.
Selective reporting	Low risk	Weight outcome outlined in methods section paper, but not reported in results section; however, detailed outcome data provided by author; "n" provided by author; raw data provided by author.
Conflict of interest	Low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. "The authors declare that there are no conflicts of interest."
Other bias	Low risk	No other potential biases are suspected.

Table S44: Risk of bias summary of Jiang et al. 2012 (study ID 3926).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	"Eggs were...given ID numbers, and evenly distributed by weight among doses"; adequate randomization is not standard protocol for these types of experiments.
Allocation concealment	Probably low risk	Eggs "given ID numbers" before allocation so there may have been allocation concealment, but not explicitly stated.
Blinding	Probably low risk	Eggs were "given ID numbers" so blinding may have been applied, but not explicitly stated.
Incomplete outcome data	Probably high risk	Hatching and mortality reported, but authors noted mortality not carefully tracked for all eggs included in outcome measurement; author reported numbers allocated.
Selective reporting	Probably low risk	Weight and length outcomes reported in results section of paper, but not pre-specified in methods section; "n" provided by author; raw data provided by author.
Selective reporting	Probably low risk	Weight and length outcomes reported in results section of paper, but not pre-specified in methods section; "n" provided by author; raw data provided by author.
Conflict of interest	Low risk	Funding source not stated, but associated persons appear to be from government and/or academia only and free of financial interests in study results. "No conflicts of interest."
Other bias	Low risk	No other potential biases are suspected.

Table S45: Risk of bias summary of Spachmo and Arukwe 2012 (study ID 3932).

Bias domain	Authors' judgment	Support for judgment
Sequence generation	Probably high risk	Randomization scheme not discussed; adequate randomization is not standard protocol for these types of experiments.
Allocation concealment	Probably high risk	Allocation concealment not discussed; adequate concealment is not standard protocol for these types of experiments.
Blinding	Probably low risk	"All cardiac and bone measurements were performed blind with respect to treatment and sampling." Unclear if applies to weight outcome.
Incomplete outcome data	High risk	Mortality not reported; only subset of treatment group analyzed for outcome; numbers allocated not reported.
Selective reporting	Low risk	Outcomes outlined in abstract/methods section of paper were reported in results section; adequate "n" reported.
Conflict of interest	Probably low risk	Associated funds and persons appear to be from government and/or academia only and free of financial interests in study results. However, no claim denying conflicts of interest was made.
Other bias	Low risk	No other potential biases are suspected.

Table S46: 95% Confidence interval estimates for mammalian fetal weight measurements, for each tested dose of PFOA (see Figure 5A).

Source [source ID]	Species	Route of Exposure	PFOA dose[a]	Lower bound	Upper bound
Yahia et al. 2010 [103]	Mouse	Gavage	1*	0.03	0.09
			5*	-0.17	-0.11
			10*	-0.44	-0.38
Fenton et al. 2009 [264]	Mouse	Gavage	0.1	-0.08	0.44
			1	0.04	0.64
			5	-0.22	0.36
Lau et al. 2006 [635][b]	Mouse	Gavage	1	-0.14	0.00
			3	-0.11	0.07
			5	-0.11	0.07
			10*	-0.18	0.04
			20*	-0.41	0.03
Staples et al. 1984 [1871]	Rat	Gavage	100	0.02	0.38
Boberg et al. 2008 [3061]	Rat	Gavage	20	-0.85	1.35
Staples et al. 1984 [1871]	Rat	Inhalation	0.1 mg/m^3	-0.23	0.03
			1 mg/m^3	-0.21	0.01
			10 mg/m^3	-0.26	0.06
			25 mg/m^3*	-0.65	-0.15

The confidence intervals are for the difference in means comparing each dose group to the control. These were calculated by us based on the information available to us (mean and standard error estimates). Asterisks indicate statistically significant ($p<0.05$) difference in response compared to control group, as reported by the study authors. The information available to us differed from that available to the study authors, and statistical tests used varied by study; therefore there is not a one-to-one correspondence with the confidence intervals as calculated by us and the p-values reported by the study authors.

[a]mg/kg BW/day, unless otherwise specified. [b]40 mg/kg BW/day group not shown because fetal weight was not measured (all offspring deceased).

Table S47: 95% Confidence interval estimates for mammalian birth weight measurements, for each tested dose of PFOA (see Figure 5B).

Source [source ID]	Species	Route of Exposure	PFOA dose[a]	Lower bound	Upper bound
Hu et al. 2010 [68]	Mouse	Drinking water	0.5*	-0.23	0.05
			1*	-0.33	-0.01
Onishchenko et al. 2011 [3610]	Mouse	Food	0.3	-0.03	0.13
Yahia et al. 2010 [103]	Mouse	Gavage	1	0.00	0.06
			5*	-0.17	-0.09
			10*	-0.51	-0.43
Hines et al. 2009 [260]	Mouse	Gavage	0.01	-0.16	0.08
			0.1	-0.19	0.03
			0.3	-0.21	0.05
			1	-0.19	0.03
			5*	-0.30	-0.02
Fenton et al. 2009 [264]	Mouse	Gavage	0.1	-0.68	-0.04
			1	-0.24	0.20
			5	-0.27	0.23
White et al. 2009 [312]	Mouse	Gavage	5	-0.33	-0.01
Abbott et al. 2007 [528]	Mouse	Gavage	0.1	-0.04	0.14
			0.3	-0.11	0.13
			0.6	-0.10	0.08
			1	-0.12	0.06
White et al. 2007 [566]	Mouse	Gavage	5*	-0.21	-0.01
Wolf et al. 2007 [571b][b]	Mouse	Gavage	5*	-0.17	0.03
			20*	-0.34	-0.10
Wolf et al. 2007 [571a][b]	Mouse	Gavage	3	-0.14	0.02
			5*	-0.21	-0.07
Lau et al. 2006 [635]	Mouse	Gavage	1	-0.10	0.12
			3	-0.10	0.16
			5	-0.14	0.00
			10	-0.31	-0.11
			20*	-0.57	-0.17
White et al. 2011 [3862]	Mouse	Gavage	1	-0.11	0.01
			5*	-0.22	-0.08
Hinderliter et al. 2005 [711]	Rat	Gavage	3	-0.30	0.78
			10	-0.17	0.73
			30	-0.01	1.01
Staples et al. 1984 [1871]	Rat	Gavage	100	-0.51	0.31
York 2002 [5122]	Rat	Gavage	1*	-0.56	-0.06
			3	-0.30	0.24
			10	-0.34	0.14
			30*	-0.80	-0.26

Source [source ID]	Species	Route of Exposure	PFOA dose[a]	Lower bound	Upper bound
Staples et al. 1984 [1871]	Rat	Inhalation	0.1 mg/m^3	-0.21	0.61
			1 mg/m^3	-0.51	0.31
			10 mg/m^3	-0.70	0.30
			25 mg/m^3*	-1.06	-0.34

The confidence intervals are for the difference in means comparing each dose group to the control. These were calculated by us based on the information available to us (mean and standard error estimates). Asterisks indicate statistically significant ($p<0.05$) difference in response compared to control group, as reported by the study authors. The information available to us differed from that available to the study authors, and statistical tests used varied by study; therefore there is not a one-to-one correspondence with the confidence intervals as calculated by us and the p-values reported by the study authors.

[a]mg/kg BW/day, unless otherwise specified. [b]Study split into 2 datasets; a) cross foster (exposure GD1-17); b) windows of sensitivity (exposure groups GD7-17, GD10-17, GD13-17, GD15-17),

Table S48: 95% Confidence interval estimates for non-mammalian weight measurements, for each tested dose of PFOA (see Figure 6A).

Source [source ID]	Species	Route of Exposure	PFOA dose	Lower bound	Upper bound
Pinkas et al. 2010 [187]	Chicken	Egg Injection	5 mg/kg egg	-2.68	2.28
			10 mg/kg egg	-2.14	3.94
O'Brien et al. 2009 [236][a]	Chicken	Egg Injection	0.01 mg/kg egg	-1.50	1.72
			0.1 mg/kg egg	-0.26	3.20
			1 mg/kg egg	-0.80	2.22
			10 mg/kg egg	0.10	3.24
Jiang et al. 2012 [3926a][b]	Chicken	Egg Injection	0.5 mg/kg egg	-0.60	7.80
			1 mg/kg egg	-1.29	5.29
			2 mg/kg egg	-2.09	6.69
Jiang et al. 2012 [3926b][b]	Chicken	Egg Injection	0.5 mg/kg egg	-3.88	1.48
			1 mg/kg egg	-2.08	2.28
			2 mg/kg egg	-2.10	5.10
Wang et al. 2010 [86]	Fruit fly	Food	100 μM*	-1.64×10^{-5}	3.64×10^{-5}
			500 μM*	-2.27×10^{-4}	-1.53×10^{-4}
Spachmo and Arukwe 2012 [3932][a]	Salmon	Egg immersion	100 μg/L water	-0.07	0.01

The confidence intervals are for the difference in means comparing each dose group to the control. These were calculated by us based on the information available to us (mean and standard error estimates). Asterisks indicate statistically significant ($p<0.05$) difference in response compared to control group, as reported by the study authors. The information available to us differed from that available to the study authors, and statistical tests used varied by study; therefore there is not a one-to-one correspondence with the confidence intervals as calculated by us and the p-values reported by the study authors.

[a]Study did not test for statistical significance. [b]Study split into 2 datasets based on time of outcome measurement a) embryonic day 19; b) 16-24 hours post hatching.

Table S49: 95% Confidence interval estimates for non-mammalian length measurements, for each tested dose of PFOA (see Figure 6B).

Source [source ID]	Species	Route of Exposure	PFOA dose	Lower bound	Upper bound
Jiang et al. 2012 [3926]	Chicken	Egg injection	0.5 mg/kg egg	-0.77	1.23
			1 mg/kg egg	-0.71	1.27
			2 mg/kg egg	-1.28	0.42
Wang et al. 2010 [86]	Fruit fly	Food	100 μM*	-0.37	-0.04
			500 μM*	-0.66	-0.24
Spachmo and Arukwe 2012 [3932][a]	Salmon	Egg immersion	100 μg/L water	-0.09	0.17
Hagenaars et al. 2011 [59]	Zebrafish	Egg immersion	15 mg/L water	-0.08	0.20
			20 mg/L water	-0.09	0.21
			30 mg/L water*	-0.22	0.06
			40 mg/L water*	-0.21	0.07
			50 mg/L water*	-0.23	0.07
			75 mg/L water*	-0.30	0.00
			100 mg/L water*	-0.42	-0.14
			250 mg/L water*	-0.69	-0.43

The confidence intervals are for the difference in means comparing each dose group to the control. These were calculated by us based on the information available to us (mean and standard error estimates). Asterisks indicate statistically significant ($p<0.05$) difference in response compared to control group, as reported by the study authors. The information available to us differed from that available to the study authors, and statistical tests used varied by study; therefore there is not a one-to-one correspondence with the confidence intervals as calculated by us and the p-values reported by the study authors.

[a]Study did not test for statistical significance.

Instructions for making risk of bias determinations

1. Sequence generation

Was the allocation sequence adequately generated?

Criteria for a judgment of 'YES' (i.e. low risk of bias):
The investigators describe a random component in the sequence generation process such as:

- Referring to a random number table;
- Using a computer random number generator;
- Coin tossing;
- Shuffling cards or envelopes;
- Throwing dice;
- Drawing of lots.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about the sequence generation process to permit a judgment of 'YES', but there is indirect evidence that suggests the sequence generation process was random, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about the sequence generation process to permit a judgment of 'NO', but there is indirect evidence that suggests a non-random component in the sequence generation process, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

The investigators describe a non-random component in the sequence generation process or that a random component was not used. Usually, the description would involve some systematic, non-random approach, for example:

- Sequence generated by date of birth;
- Sequence generated by some rule based on date (or day) of arrival at facility;
- Sequence generated by some rule based on record number.

Other non-random approaches happen much less frequently than the systematic approaches mentioned above and tend to be obvious. They usually involve judgment or some method of non-random categorization of animals, for example:

- Allocation by judgment of the investigator;
- Allocation by availability of the intervention.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that sequence generation is not an element of study design capable of introducing risk of bias in the study.

2. Allocation concealment

Was allocation adequately concealed?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

Investigators could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation:

- Sequentially numbered treatment containers of identical appearance to control; or
- Sequentially numbered prepared route of administration (e.g., pre-prepared water dosed with chemical) of identical appearance; or
- Sequentially numbered, opaque, sealed envelopes.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about allocation concealment to permit a judgment of 'YES', but there is indirect evidence that suggests the allocation was adequately concealed, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about allocation concealment to permit a judgment of 'NO', but there is indirect evidence that suggests the allocation was not adequately concealed, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

Investigators handling experimental animals could possibly foresee assignments and thus introduce selection bias, such as allocation based on:

- Using an open random allocation schedule (e.g. a list of random numbers); or
- Assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered); or
- Alternation or rotation; or
- Non-random and known criteria, such as date of birth; or
- Record number; or
- Any other explicitly unconcealed procedure.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that allocation concealment is not an element of study design capable of introducing risk of bias in the study.

3. Blinding of personnel and outcome assessors

Was knowledge of the allocated interventions adequately prevented during the study?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

Any one of the following:

- No blinding, but the review authors judge that the outcome and the outcome measurement are not likely to be influenced by lack of blinding; or
- Blinding of key study personnel ensured, and unlikely that the blinding could have been broken; or
- Some key study personnel were not blinded, but outcome assessment was blinded and the non-blinding of others unlikely to introduce bias.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about blinding to permit a judgment of 'YES', but there is indirect evidence that suggests the study was adequately blinded, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about blinding to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not adequately blinded, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

Any one of the following:

- No blinding or incomplete blinding, and the outcome or outcome measurement is likely to be influenced by lack of blinding; or
- Blinding of key study personnel attempted, but likely that the blinding could have been broken; or
- Some key study personnel were not blinded, and the non-blinding of others likely to introduce bias.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that blinding is not an element of study design capable of introducing risk of bias in the study.

4. Incomplete outcome data

Were incomplete outcome data adequately addressed?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

The number of animals assessed for outcome of interest is reported and data is provided indicating adequate follow up of all treated animals. Additional information provided by authors should be considered when making risk of bias judgments about incomplete outcome data. Additionally, any one of the following:

- No missing outcome data; or
- Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias); or
- Missing outcome data is provided and is balanced in numbers across intervention groups, with similar reasons for missing data across groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a biologically relevant impact on the intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a biologically relevant impact on observed effect size; or
- Missing data have been imputed using appropriate statistical methods.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about incomplete outcome data to permit a judgment of 'YES', but there is indirect evidence that suggests incomplete outcome data were adequately addressed, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about incomplete outcome data to permit a judgment of 'NO', but there is indirect evidence that suggests incomplete outcome data was not adequately addressed, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

The number of animals allocated not reported and no data is provided to indicate that there was adequate follow up of all treated animals. Additionally, any one of the following:

- Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups; or
- For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce biologically relevant bias in intervention effect estimate; or
- For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce biologically relevant bias in observed effect size; or
- 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization; or
- Potentially inappropriate application of simple imputation.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that incomplete outcome data is not capable of introducing risk of bias in the study.

5. Selective outcome reporting

Are reports of the study free of suggestion of selective outcome reporting?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

All of the study's pre-specified (primary and secondary) outcomes outlined in the methods, abstract, and/or introduction that are of interest in the review have been reported in the pre-specified way, including the number of animals analyzed for outcomes of interest. Additional information provided by authors should be considered when making risk of bias judgments for selective outcome reporting.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information about selective outcome reporting to permit a judgment of 'YES', but there is indirect evidence that suggests the study was free of selective reporting, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information about selective outcome reporting to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not free of selective reporting, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

The study was not free of selective reporting. The following should be considered:

- Authors did not report numbers analyzed for outcomes of interest; or
- Not all of the study's pre-specified primary outcomes (as outlined in the protocol, title, abstract, and/or introduction) that are of interest in the review have been reported; or
- One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); or
- One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect); or
- One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis; or
- The study report fails to include results for a key outcome that would be expected to have been reported for such a study.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that selective outcome reporting is not capable of introducing risk of bias in the study.

6. Other potential threats to validity

Was the study apparently free of other problems that could put it at a risk of bias?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

The study appears to be free of other sources of bias.

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information to permit a judgment of 'YES', but there is indirect evidence that suggests the study was free of other threats to validity, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not free of other threats to validity, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

There is at least one important risk of bias. For example, the study:

- Had a potential source of bias related to the specific study design used;
- Stopped early due to some data-dependent process (including a formal-stopping rule);
- Had extreme baseline imbalance (improper control group);
- Has been claimed to have been fraudulent;
- The conduct of the study is affected by interim results (e.g. recruiting additional animals from a subgroup showing more benefit);
- There is deviation from the study protocol in a way that does not reflect typical practice (e.g. post hoc stepping-up of doses to exaggerated levels);
- There is pre-randomization administration of an intervention that could enhance or diminish the effect of a subsequent, randomized, intervention; inappropriate administration of an intervention (or co-intervention);
- Occurrence of 'null bias' due to interventions being insufficiently well delivered or overly wide inclusion criteria for animals;
- An insensitive instrument is used to measure outcomes (which can lead to under-estimation of both beneficial and harmful effects);
- Selective reporting of subgroups;
- Had some other problem.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that other potential threats to validity are not capable of introducing risk of bias in the study.

7. Conflict of interest

Was the study free of support from a company, study author, or other entity having a financial interest in any of the treatments studied?

Criteria for a judgment of 'YES' (i.e. low risk of bias):

The study did not receive support from a company, study author, or other entity having a financial interest in the outcome of the study. A conflict of interest statement is provided to indicate the authors have no financial interest and there is evidence of the entities not having a financial interest. Examples of this evidence include the following:

- Funding source is limited to government, non-profit organizations, or academic grants funded by government, foundations and/or non-profit organizations;
- Chemicals or other treatment used in study were purchased from a supplier;
- Company affiliated staff are not mentioned in the acknowledgements section;
- Authors were not employees of a company with a financial interest in the outcome of the study;
- Company with a financial interest in the outcome of the study was not involved in the design, conduct, analysis, or reporting of the study and authors had complete access to the data;
- Study authors make a claim denying conflicts of interest;
- Study authors are unaffiliated with companies with financial interest, and there is no reason to believe a conflict of interest exists;
- All study authors are affiliated with a government agency (are prohibited from involvement in projects for which there is a conflict of interest or an appearance of conflict of interest).

Criteria for the judgment of 'PROBABLY YES' (i.e. probably low risk of bias):

There is insufficient information to permit a judgment of 'YES', for example there is no conflict of interest statement denying financial interests, but there is evidence that suggests the study was free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of 'YES'.

Criteria for the judgment of 'PROBABLY NO' (i.e. probably high risk of bias):

There is insufficient information to permit a judgment of 'NO', but there is indirect evidence that suggests the study was not free of support from a company, study author, or other entity having a financial interest in the outcome of the study, as described by the criteria for a judgment of 'NO'.

Criteria for the judgment of 'NO' (i.e. high risk of bias):

The study received support from a company, study author, or other entity having a financial interest in the outcome of the study. Examples of support include:

- Research funds;
- Writing services;
- Author/staff from study was employee or otherwise affiliated with company with financial interest;

- Company limited author access to the data;
- Company was involved in the design, conduct, analysis, or reporting of the study;
- Study authors claim a conflict of interest.

Criteria for the judgment of 'NOT APPLICABLE' (risk of bias domain is not applicable to study):

There is evidence that conflicts of interest are not capable of introducing risk of bias in the study.

References

Abbott BD, Wolf CJ, Schmid JE, Das KP, Zehr RD, Helfant L, et al. 2007. Perfluorooctanoic acid induced developmental toxicity in the mouse is dependent on expression of peroxisome proliferator activated receptor-alpha. Toxicol Sci 98:571-581.

ATSDR (Agency for Toxic Substances & Disease Registry). 2012. Interaction Profiles for Toxic Substances. Available: http://www.atsdr.cdc.gov/interactionprofiles/index.asp [accessed 6 February 2012].

ATSDR (Agency for Toxic Substances & Disease Registry). 2012. Toxic Substances Portal. Available: http://www.atsdr.cdc.gov/toxprofiles/index.asp [accessed 6 February 2012].

Boberg J, Metzdorff S, Wortziger R, Axelstad M, Brokken L, Vinggaard AM, et al. 2008. Impact of diisobutyl phthalate and other PPAR agonists on steroidogenesis and plasma insulin and leptin levels in fetal rats. Toxicology 250:75-81.

Butenhoff JL, Kennedy GL, Jr., Frame SR, O'Connor JC, York RG. 2004. The reproductive toxicology of ammonium perfluorooctanoate (APFO) in the rat. Toxicology 196:95-116.

CalEPA (California Environmental Protection Agency). 2012. Office of Environmental Health Hazard Assessment Risk Assessment. Available: http://www.oehha.ca.gov/risk.html [accessed 6 February 2012].

European Chemicals Agency. 2012. Homepage. Available: http://echa.europa.eu/web/guest/home [accessed 6 February 2012].

Fenton SE, Reiner JL, Nakayama SF, Delinsky AD, Stanko JP, Hines EP, et al. 2009. Analysis of PFOA in dosed CD-1 mice. Part 2. Disposition of PFOA in tissues and fluids from pregnant and lactating mice and their pups. Reprod Toxicol 27:365-372.

Hagenaars A, Vergauwen L, De Coen W, Knapen D. 2011. Structure-activity relationship assessment of four perfluorinated chemicals using a prolonged zebrafish early life stage test. Chemosphere 82:764-772.

Health Canada. 2012. First Priority Substances List (PSL1) Assessments. Available: http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl1-lsp1/index-eng.php [accessed 3 February 2012].

Health Canada. 2012. Second Priority Substances List (PSL2) Assessments. Available: http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/psl2-lsp2/index-eng.php [accessed 3 February 2012].

Hinderliter PM, Mylchreest E, Gannon SA, Butenhoff JL, Kennedy GL Jr. 2005. Perfluorooctanoate: Placental and lactational transport pharmacokinetics in rats. Toxicology 211:139-148.

Hines EP, White SS, Stanko JP, Gibbs-Flournoy EA, Lau C, Fenton SE. 2009. Phenotypic dichotomy following developmental exposure to perfluorooctanoic acid (PFOA) in female CD-1 mice: Low doses induce elevated serum leptin and insulin, and overweight in mid-life. Mol Cell Endocrinol 304:97-105.

Hooijmans CR, Tillema A, Leenaars M, Ritskes-Hoitinga M. 2010b. Enhancing search efficiency by means of a search filter for finding all studies on animal experimentation in PubMed. Lab Anim 44:170-175.

Hu Q, Strynar MJ, DeWitt JC. 2010. Are developmentally exposed C57BL/6 mice insensitive to suppression of TDAR by PFOA? J Immunotoxicol 7:344-349.

IARC (International Agency for Research on Cancer). 2012. Monographs on the Evaluation of Carcinogenic Risks to Humans. Available: http://monographs.iarc.fr/ENG/Monographs/PDFs/index.php [accessed 3 February 2012].

ILSI (International Life Sciences Institute). 2012. Homepage. Available: http://www.ilsi.org/Pages/HomePage.aspx [accessed 3 February 2012].

IPCS (International Programme on Chemical Safety). 2012. INCHEM. Available: http://www.inchem.org/ [accessed 3 February 2012].

Jiang Q, Lust RM, Strynar MJ, Dagnino S, DeWitt JC. 2012. Perflurooctanoic acid induces developmental cardiotoxicity in chicken embryos and hatchlings. Toxicology 293:97-106.

Lau C, Thibodeaux JR, Hanson RG, Narotsky MG, Rogers JM, Lindstrom AB, et al. 2006. Effects of perfluorooctanoic acid exposure during pregnancy in the mouse. Toxicol Sci 90:510-518.

National Center for Biotechnology Information. 2012. PubMed. Available: http://www.ncbi.nlm.nih.gov/pubmed/ [accessed 3 February 2012].

National Library of Medicine. 2012. Chemical Carcinogenesis Research Information System (CCRIS). Available: http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS [accessed 6 February 2012].

National Library of Medicine. 2012. ChemIDplus Advanced. Available: http://chem.sis.nlm.nih.gov/chemidplus/ [accessed 6 February 2012].

National Library of Medicine. 2012. Developmental and Reproductive Toxicology Database (DART). Available: http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?DARTETIC [accessed 6 February 2012].

National Library of Medicine. 2012. Environmental Mutagen Information Center (EMIC). Available: http://toxnet.nlm.nih.gov/ [accessed 6 February 2012].

National Library of Medicine. 2012. Genetic Toxicology Data Bank (GENE-TOX). Available: http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?GENETOX [accessed 6 February 2012].

National Library of Medicine. 2012. Hazardous Substances Data Bank (HSDB). Available: http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB [accessed 6 February 2012].

National Library of Medicine. 2012. Toxicology Data Network (TOXNET). Available: http://toxnet.nlm.nih.gov/ [accessed 6 February 2012].

National Library of Medicine. 2012. Toxicology Literature Online (TOXLINE). Available: http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLINE [accessed 6 February 2012].

NIOSH (National Institute for Occupational Safety and Health). 2012. NIOSHTIC-2. Available: https://www2a.cdc.gov/nioshtic-2/advsearch2.asp [accessed 6 February 2012].

NIOSH (National Institute for Occupational Safety and Health). 2012. Registry of Toxic Effects of Chemical Substances (RTECS). Available: http://www.cdc.gov/niosh/rtecs/default.html [accessed 6 February 2012].

NTP (National Toxicology Program). 2012. Management Status Report. Available: http://ntp.niehs.nih.gov/go/MSR-index [accessed 6 February 2012].

NTP (National Toxicology Program). 2012. Program Results and Status Search. Available: http://ntp.niehs.nih.gov/?objectid=720160DB-BDB7-CEBA-F7CC2DE0A230C920 [accessed 6 February 2012].

NTP (National Toxicology Program). 2012. Program Report on Carcinogens. Available: http://ntp.niehs.nih.gov/?objectid=72016262-BDB7-CEBA-FA60E922B18C2540 [accessed 6 February 2012].

O'Brien JM, Crump D, Mundy LJ, Chu S, McLaren KK, Vongphachan V, et al. 2009. Pipping success and liver mRNA expression in chicken embryos exposed in ovo to C8 and C11 perfluorinated carboxylic acids and C10 perfluorinated sulfonate. Toxicol Lett 190:134-139.

Onishchenko N, Fischer C, Wan Ibrahim WN, Negri S, Spulber S, Cottica D, et al. 2011. Prenatal exposure to PFOS or PFOA alters motor function in mice in a sex-related manner. Neurotox Res 19:452-461.

Pinkas A, Slotkin TA, Brick-Turin Y, Van der Zee EA, Yanai J. 2010. Neurobehavioral teratogenicity of perfluorinated alkyls in an avian model. Neurotoxicol Teratol 32:182-186.

Royal Society of Chemistry. 2012. ChemSpider. Available: http://www.chemspider.com/ [accessed 6 February 2012].

Spachmo B, Arukwe A. 2012. Endocrine and developmental effects in Atlantic salmon (Salmo salar) exposed to perfluorooctane sulfonic or perfluorooctane carboxylic acids. Aquat Toxicol 108:112-124.

Staples RE, Burgess BA, Kerns WD. 1984. The embryo-fetal toxicity and teratogenic potential of ammonium perfluorooctanoate (APFO) in the rat. Fundam Appl Toxicol 4:429-440.

Thompson Reuters. 2012. Web of Science. Available: http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=3BIcWf14OmZnOOyxRJ9&preferencesSaved= [accessed 3 February 2012].

Toxicology Excellence for Risk Assessment. 2012. Homepage. Available: http://www.tera.org/ [accessed 3 February 2012].

Toxicology Excellence for Risk Assessment. 2012. International Toxicity Estimates for Risk (ITER). Available: http://iter.ctcnet.net/publicurl/pub_search_list.cfm [accessed 3 February 2012].

US EPA (U.S. Environmental Protection Agency). 2012. Acute Exposure Guideline Levels Chemicals. Available: http://www.epa.gov/oppt/aegl/pubs/chemlist.htm [accessed 6 February 2012].

US EPA (U.S. Environmental Protection Agency). 2012. EPA Science Inventory. Available: http://cfpub.epa.gov/si/ [accessed 6 February 2012].

US EPA (U.S. Environmental Protection Agency). 2012. Health & Environmental Research Online (HERO). Available: http://hero.epa.gov/ [accessed 6 February 2012].

US EPA (U.S. Environmental Protection Agency). 2012. Integrated Risk Information System (IRIS). Available: http://www.epa.gov/iris/ [accessed 6 February 2012].

US EPA (U.S. Environmental Protection Agency). 2012. National Environmental Publications Internet Site (NEPIS). Available: http://nepis.epa.gov [accessed 6 February 2012].

US EPA (U.S. Environmental Protection Agency). 2012. National Service Center for Environmental Publications (NSCEP). Available: http://www.epa.gov/ncepihom/ [accessed 6 February 2012].

US EPA (U.S. Environmental Protection Agency). 2012. Substance Registry Services. Available: http://www.epa.gov/srs/ [accessed 6 February 2012].

US EPA (U.S. Environmental Protection Agency). 2012. Toxic Substances Control Act Test Submissions (TSCATS). Available: http://www.ntis.gov/products/ots.aspx [accessed 6 February 2012].

Wang J, Li Y, Liu Y, Zhang H, Dai J. 2010. Disturbance of perfluorooctanoic acid on development and behavior in Drosophila larvae. Environ Toxicol Chem 29:2117-2122.

White SS, Calafat AM, Kuklenyik Z, Villanueva L, Zehr RD, Helfant L, et al. 2007. Gestational PFOA exposure of mice is associated with altered mammary gland development in dams and female offspring. Toxicol Sci 96:133-144.

White SS, Kato K, Jia LT, Basden BJ, Calafat AM, Hines EP, et al. 2009. Effects of perfluorooctanoic acid on mouse mammary gland development and differentiation resulting from cross-foster and restricted gestational exposures. Reprod Toxicol 27:289-298.

White SS, Stanko JP, Kato K, Calafat AM, Hines EP, Fenton SE. 2011. Gestational and chronic low-dose PFOA exposures and mammary gland growth and differentiation in three generations of CD-1 mice. Environ Health Perspect 119:1070-1076.

WHO (World Health Organization). 2012. Concise International Chemical Assessment Documents. Available: http://www.who.int/ipcs/publications/cicad/en/index.html [accessed 6 February 2012].

WHO (World Health Organization). 2012. Environmental Health Criteria. Available: http://www.who.int/ipcs/publications/ehc/en/index.html [accessed 6 February 2012].

Wolf CJ, Fenton SE, Schmid JE, Calafat AM, Kuklenyik Z, Bryant XA, et al. 2007. Developmental toxicity of perfluorooctanoic acid in the CD-1 mouse after cross-foster and restricted gestational exposures. Toxicol Sci 95:462-473.

Yahia D, El-Nasser MA, Abedel-Latif M, Tsukuba C, Yoshida M, Sato I, et al. 2010. Effects of perfluorooctanoic acid (PFOA) exposure to pregnant mice on reproduction. J Toxicol Sci 35:527-533.

York R. 2002. Oral (gavage) two-generation (one litter per generation) reproduction study of ammonium perfluorooctanoate (APFO) in rats. Report prepared for 3M, St. Paul, MN. Sponsor's Study No T-6889.6. Reviewed in US EPA AR226-1092. Horsham, PA:Argus Research.